

Pursing Data Analytics

The Complete Guide

A handbook for Chemical Engineering & Applied Chemistry Graduate Students

Contents

Introduction	2
What Courses Should You Take?	3
CHE courses where data science could be applied	3
Data science courses for beginner students	3
Data science courses for advanced students	3
MEng Projects in Data Science	6
Data Science Organizations on Campus	6

Introduction

Do you have a passion for data analytics or machine learning? If so, this guide will help you to answer the following questions:

- Where can I find more analytics or data science courses?
- How can I do an MEng project in analytics or data science?
- How do I get more involved in the University's data science community?

What Courses Should You Take?

CHE courses where data science could be applied

- [CHE1107H](#) - Applied Mathematics
- [CHE1141H](#) – Advanced Chemical Reaction Engineering
- [CHE1142H](#) - Applied Chemical Thermodynamics
- [CHE1147H](#) - Data Mining in Engineering
- [CHE1148H](#) - Data Process Analytics
- [CHE1434H](#) - Six Sigma for Chemical Processes
- [CHE1436H](#) - Risk Assessment for Chemical Process Safety
- [CHE1471H](#) - Modelling in Biological and Chemical Systems
- [CHE507H1](#) - Data-based Modelling for Prediction and Control

Data science courses for beginner students

The [Analytics Emphasis](#) has many courses related to data science that you can pursue from different departments in the University. There are good courses for beginners in this list.

Data science courses for advanced students

If you are looking for advanced options and want to explore some courses from the [Department of Computer Science](#) or the [Faculty of Information Studies](#), look at these:

- **CSC2511HS/CSC401H1 — Natural Language Computing**
Introduction to techniques involving natural language and speech in applications such as intelligent Web searching; speech recognition and synthesis; and multi-lingual systems including machine translation. N-grams, POS-tagging, semantic distance metrics, neural models of language, corpus analysis. Python and other software. Prerequisite: CSC207H1/ CSC209H1; STA247H1/ STA255H1/ STA257H1
- **CSC2515H – Introduction to Machine Learning**
The students will learn about ML problems (supervised, unsupervised, and reinforcement learning), models (linear and nonlinear, including neural networks), loss functions (squared error, cross entropy, hinge, exponential), bias and variance trade off, ensemble methods (bagging and boosting), optimization techniques in ML, probabilistic viewpoint of ML, etc. Prerequisite: This is a relatively math-heavy course. The students need to be comfortable with the following topics: Probability Theory, Multivariate Calculus, Linear Algebra, Programming (preferably in Python). MAT235Y1/ MAT237Y1/ MAT257Y1; MAT221H1/ MAT223H1/ MAT240H1; STA247H1/ STA255H1/ STA257H1
- **CSC2516HS/CSC413H1 — Neural Networks and Deep Learning**
It is very hard to hand-design programs to solve many real-world problems, e.g. distinguishing images of cats vs. dogs. Machine learning algorithms allow computers to learn from example data and produce a program that does the job. Neural networks are a class of

machine learning algorithms originally inspired by the brain, but which have recently have seen a lot of success at practical applications. They're at the heart of production systems at companies like Google and Facebook for image processing, speech-to-text, and language understanding. This course gives an overview of both the foundational ideas and the recent advances in neural net algorithms.

- **CSC2518HS — Spoken Language Processing**
 An introduction to working with speech in natural language processing systems. Topics include articulatory and acoustic phonetics, prosody and information structure, introduction to digital signal processing of speech, automated speech recognition, text-to-speech synthesis, language models, dialogue modelling and dialogue systems.
- **CSC2541HF — Topics in Machine Learning: Machine Learning for Health**
 This course will give a broad overview of machine learning for health. We begin with an overview of what makes healthcare unique, and then explore machine learning methods for clinical and healthcare applications through recent papers. We discuss the recent successes of graphical models, deep learning, time series analysis, and transfer learning in the context of health.
- **CSC2541HS — Topics in Machine Learning: Neural Network Training Dynamics**
 Neural networks have driven remarkable progress across a range of applications but understanding why they work can still be a challenge. This course covers a set of mathematical tools useful for analysing and understanding neural net training phenomena, and for diagnosing failures. Topics include linearized network approximations, stochastic optimization, wide/deep network limits, continuous time limits, and linear systems. We'll apply these tools to understanding single level, minmax, and bilevel optimization of neural nets.
- **CSC2547HS — Current Algorithms and Techniques in Machine Learning: Machine Learning for Vision as Inverse Graphics**
 This is an advanced graduate course in machine learning. It includes some lectures, especially at the beginning, but is primarily a seminar course in which students will read and present papers from the literature. There will also be a course project and possibly some assignments. The goal is to bring students to the state of the art in this exciting field. Tentative topics include generative and discriminative models for vision, convolutional and deconvolutional neural nets, variational inference and autoencoders, capsule networks, group symmetries and equivariance, visual attention mechanisms, vision transformers, differentiable renderers, and applications. Mathematical maturity is required.
- **CSC2559HF — Trustworthy Machine Learning**
 The deployment of machine learning in real-world systems calls for a set of complementary technologies that will ensure that machine learning is trustworthy. Here, the notion of trust is used in its broad meaning: the course covers different topics in emerging research areas related to the broader study of security and privacy in machine learning. Students will learn about attacks against computer systems leveraging machine learning, as well as defense techniques to mitigate such attacks. The course assumes students already have a basic understanding of machine learning. Students will familiarize themselves with the emerging

body of literature from different research communities investigating these questions. The class is designed to help students explore new research directions and applications.

- **ETH1000HY** — Ethics of Artificial Intelligence in Context (This year-long, half-credit graduate course exposes students to advanced methods employed in the analysis of ethical issues related to the production, dissemination, and application of AI in a variety of contexts).
- [INF1340H](#) - Programming for Data Science
- [INF1343H](#) - Data Modelling and Database Design
- [INF1344H](#) - Introduction to Statistics for Data Science
- [INF2167H](#) - R for Data Science
- [INF2178H](#) - Experimental Design for Data Science
- [INF2179H](#) - Machine Learning with Applications in Python
- [INF2190H](#) - Data Analytics: Introduction, Methods, and Practical Approaches
- [INF2210H](#) - Human Values in Data Sciences
- [INF2202H](#) - Data Governance in a World of Big Data
- [INF2203H](#) - Storytelling with Data

MEng Projects in Data Science

CHE faculty members who do data science and/or machine learning related MEng projects:

1. [Greg J. Evans](#) (Professor)
2. [Daniela Galatro](#) (Assistant Professor)
3. [Yuri A. Lawryshyn](#) (Professor)
4. [Radhakrishnan Mahadevan](#) (Professor)

Data Science Organizations on Campus

To connect with the University's data science community, check out these organizations:

[Centre for Analytics and Artificial Intelligence Engineering](#)

The Centre for Analytics and AI Engineering (CARTE) is the hub for collaborations in Analytics/AI/ML within the Faculty of Applied Science and Engineering. We provide broad support to faculty members, students, and industry partners. In particular, we would like to highlight the following services for students:

- Industry speaker seminar series
- AI/ML research drop-in clinic
- Analytics/AI/ML student CV bank: Submit your CVs to get potential connections with Professors who are looking for students to take up projects!
- MITACS internship opportunities in Analytics/AI/ML
- Industry Job Board: We announce employment opportunities with our partners here

[Data Science & Statistics Society](#) (UTSC organization, does not specify if it is only for undergrads)

The purpose of The Data Science and Statistics Society (The Data Science Cube) is to create a platform for peer-mentorship, career and professional skills development among students interested in Data Science and Statistics.

[Humans Learning Machine Learning \(HLML\)](#) (Physics group – but does not mention that they only take physics students, or are only for undergrad students)

They are a group of students at the University of Toronto who are doing research in computational and physical sciences using the powers of machine learning, statistics, and high-performance computing to tackle interesting and fun problems.