Disclaimer

If a student turns on their microphone or camera or uses the public chat feature, this constitutes consent for the student's video image or sound audio to be uploaded with the office hour or tutorial on university approved platforms such as D2L. If the student wishes to ensure that their questions/faces/voices are not recorded in the video, they should instead use private chat to ask questions.

ENEL 645 – Data Mining & Machine Learning

Overview of the course

Roberto Souza
Assistant Professor
Electrical and Computer Engineering
Schulich School of Engineering

January 2022

UNIVERSITY OF CALGARY

Instructor and TAs

Instructor:

- Roberto Souza – roberto.medeirosdeso@ucalgary.ca

TAs:

- Mike Lasby – mklasby@ucalgary.ca
- Youssef Beauferris – youssef.beauferris@ucalgary.ca

- Please avoid contacting the instructor and TAs directly by email unless it is an issue specific to a single student or to a project team.

Course Delivery

- Synchronous and in-person (when safe to do so - January is expected to be delivered online)
  - MWF – 9:00 am to 9:50 am
  - Room TI 160

- One 1-hour office hour per week (more if necessary)
  - Starts on the 2nd week of class
  - Dates and times to be confirmed

- It is a large class, but **don't be a stranger! Come to the office hours with your questions.**

Course Syllabus

Types of data mining: **classification, clustering, association, prediction. Processes: data preparation, model building**. Techniques: decision tree, **neural network**, evolutionary computing, Bayesian network. Applications: multi-media, text and web mining.

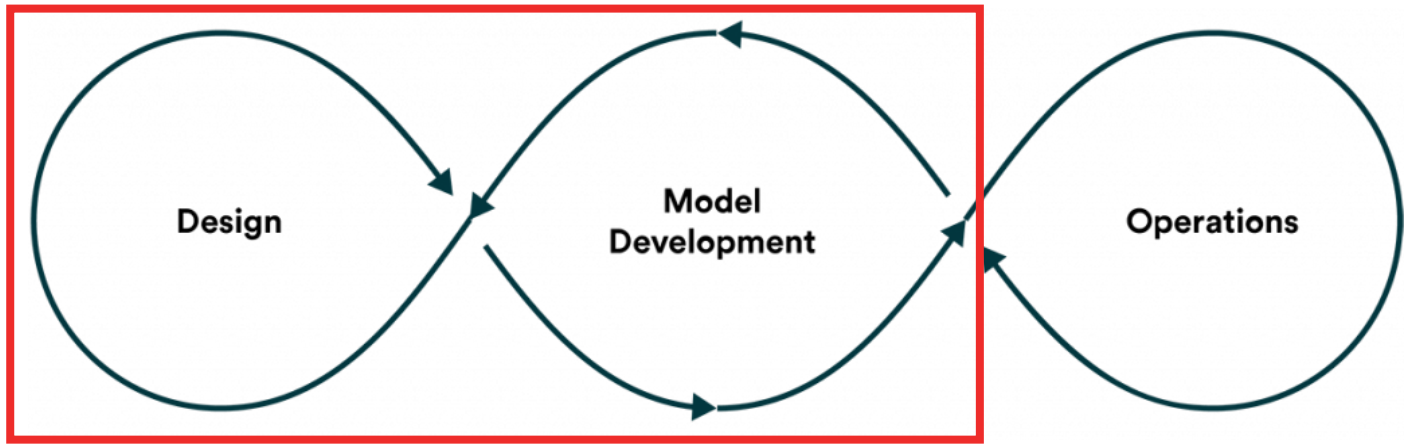
Course Syllabus (main topics)

- Python Bootcamp and machine learning concepts
- Data preparation and pre-processing
- Regularization techniques
- Traditional machine learning models (Decision Trees, Random Forests, …)
- Neural Networks
- Transfer Learning and Domain Adaptation
- Generative adversarial networks
- Self-supervised learning
- Automated Machine Learning

Learning Outcomes

1. Design and develop data mining and machine learning solutions for relevant problems
2. Select appropriate experimental setups and metrics for evaluating machine learning models
3. Select appropriate machine learning models for different types of problems
4. Have a comprehensive overview of current trends in machine learning
5. Acquire hands-on experience with machine learning programming frameworks

What this course is about?



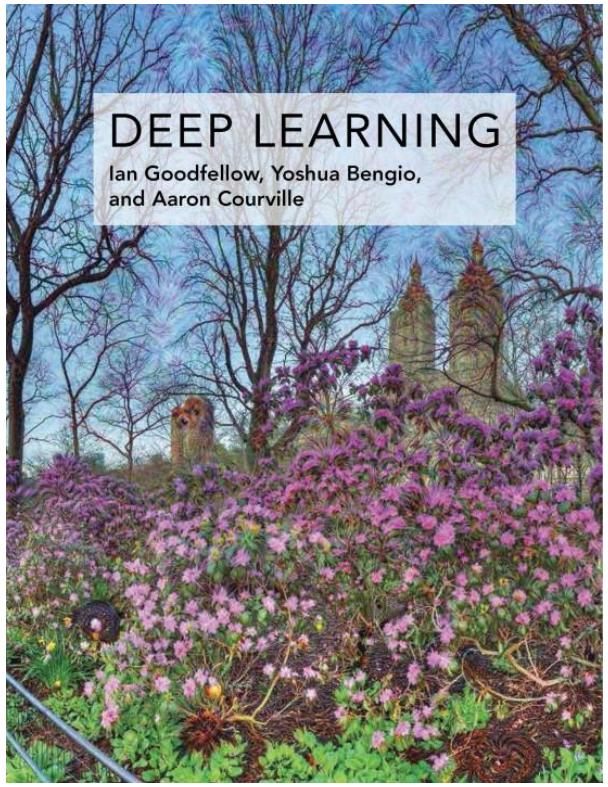
- Model deployment
- CI/CD pipelines
- Monitoring & triggering

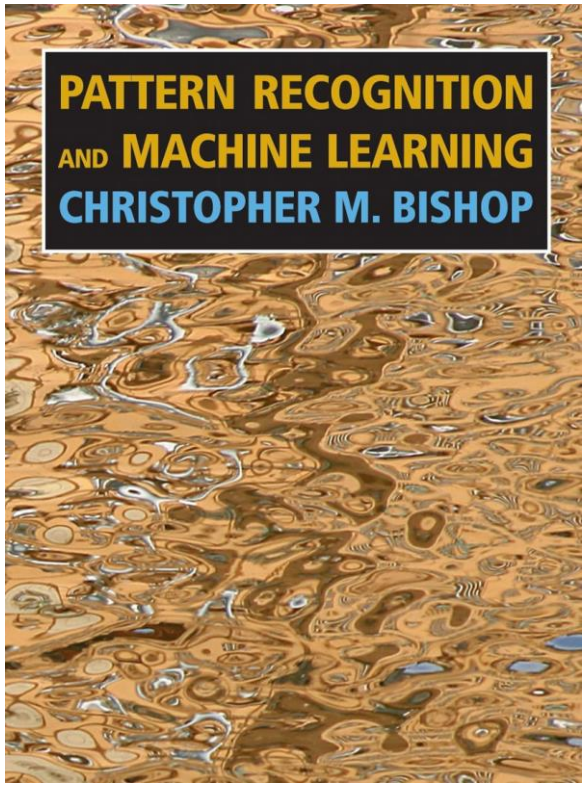
This course is about designing and developing machine learning models to achieve the best quantitative metrics* to the problems being modelled.

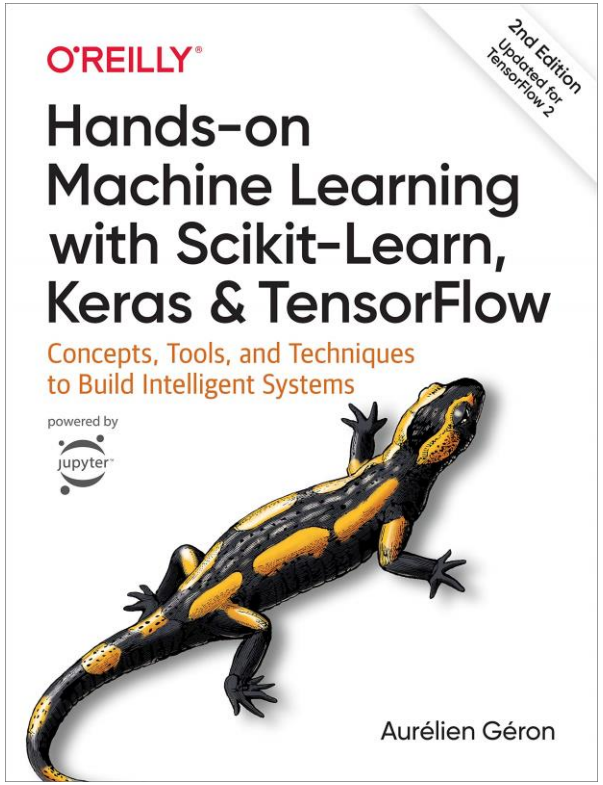
*Quantitative results are determined on problem-by-problem basis.

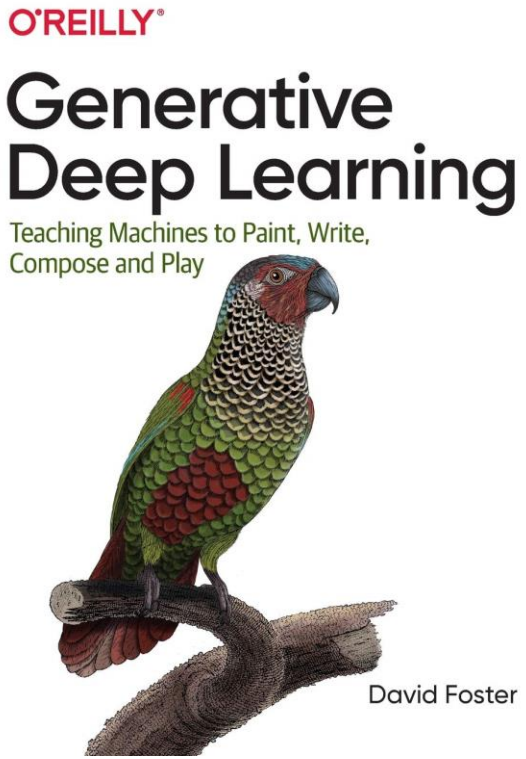
Textbook

- No mandatory textbook for this course









Class GitHub repo: https://github.com/rmsouza01/ENEL645

Course Assessment

Component	Learning Outcomes	Weight
Assignments	1, 2, 3, 4, 5	30%
Quizzes	1, 4, 6	20%
Final Project	1, 2, 3, 4, 5, 6	50%

- The lowest quiz grade will be dropped

Quizzes (20%)

- 3 quizzes – 2 highest grades are kept
- After starting, you have 120 minutes to finish the quiz
- Quizzes are individual
- **Quiz 01**
  - Available on 02 February at 9 am
  - Due on 04 February at noon
- **Quiz 02**
  - Available on 02 March at 9 am
  - Due on 04 March at noon
- **Quiz 03**
  - Available on 01 April at 9 am
  - Due on 04 April at noon

Assignments (30%)

- Team-based – maximum 6 people per team
- **Assignment 01 (choose one):**
  - Python/NumPy programming
  - Proposing an image-based garbage classification system or a problem of interest to your research
  - **Due:** Noon - 31 January 2022 | **Delivery method:** GitHub repository or written report
- **Assignment 02**
  - Summarize a pre-selected paper - teams will be given options
  - **Due:** Noon -08 March | **Delivery method:** Pre-recorded presentation on D2L dropbox
- **Assignment 03 (choose one)**
  - Building a classification model
  - Implementing a domain adaptation method for an image classification problem
  - Implementing a signal denoising model
  - Implementing a generative adversarial model
  - **Due:** Noon - 01 April 2022 | **Delivery method:** GitHub repository

Final Project (50%)

- 8-page report + 1 additional page only with references (if necessary)
- Report template
  - Overleaf - please make a copy for your team.
  - Microsoft word
- Report due date: **04 April at 9 am**
- 8-minute presentation + 2 minutes for questions – 4, 6, 8, 11 April

Grades

Letter Grade	Total Mark (T)
A+	T≥95%
A	90% ≤ T < 95%
A-	85% ≤ T < 90%
B+	80% ≤ T < 85%
B	75% ≤ T < 80%
B-	70% ≤ T < 75%
C+	65% ≤ T < 70%
C	60% ≤ T < 65%
C-	55% ≤ T < 60%
D+	50% ≤ T < 55%
D	45% ≤ T < 50%
F	T < 45%

The Programming Environment (Part 1)

https://colab.research.google.com/

https://jupyter.org/

https://github.com/rmsouza01/ENEL645

https://www.overleaf.com/project

https://rcs.ucalgary.ca/index.php/RCS_Home_Page

The Programming Environment (Part 2)

- Python 3
- Python libraries:
  - NumPy
  - Matplotlib
  - Pandas
  - Scikit-learn
  - Tensorflow (version ≥ 2.0)

- Please have your programming environment in your computer or on Google Colab set up asap

Deep Learning Framework

Calibration Survey

- Please complete the following survey:

https://forms.gle/eeFyDoQsSuTwVcwA7

- It takes less than a minute to do it!

I hope you enjoy the class ☺

Questions?