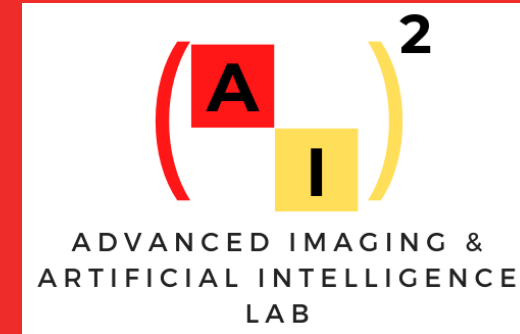


ENEL 645 – Data Mining & Machine Learning

Overview of the course

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

W2024



@lab_ai2



UNIVERSITY OF
CALGARY

Instructor and TAs

Instructors:

- Roberto Souza – roberto.souza2@ucalgary.ca
- Natalia Dubljevic – natalia.dubljevic@ucalgary.ca

TAs:

- Mahsa Dibaji – seyedemahsa.dibaji@ucalgary.ca
- Peyman Tahghighi – peyman.tahghighi@ucalgary.ca
- Please avoid contacting the instructor and TAs directly by email unless it is an issue specific to a grade produced by the TAs.

Course Delivery

- Synchronous and in-person
 - MWF – 9:00 am to 9:50 am
 - Room SA 104
- Use the D2L discussion board for 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.

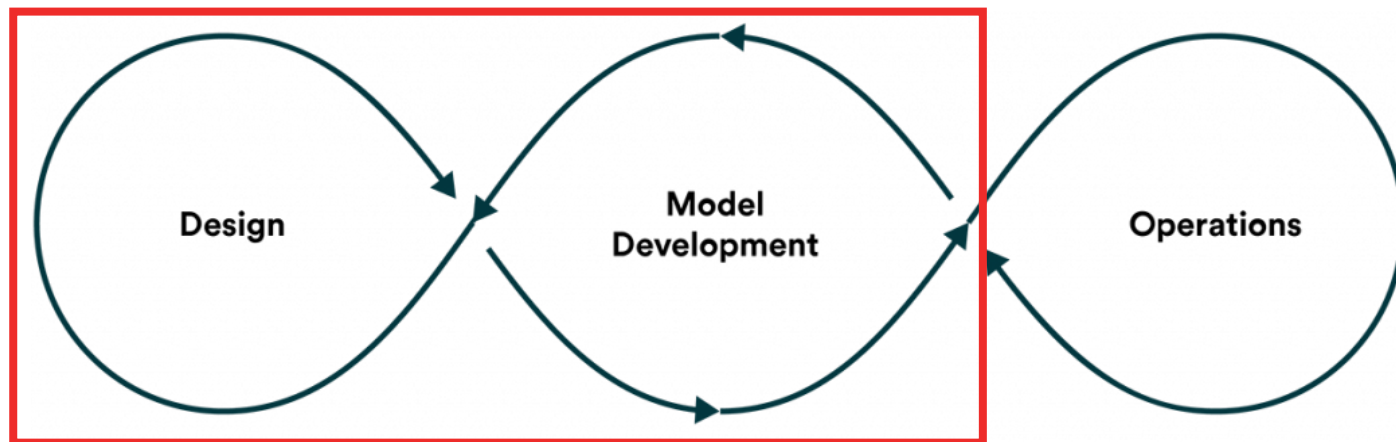
Course Syllabus (main topics)

- 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 models
- Self-supervised learning
- Physics informed neural networks

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?

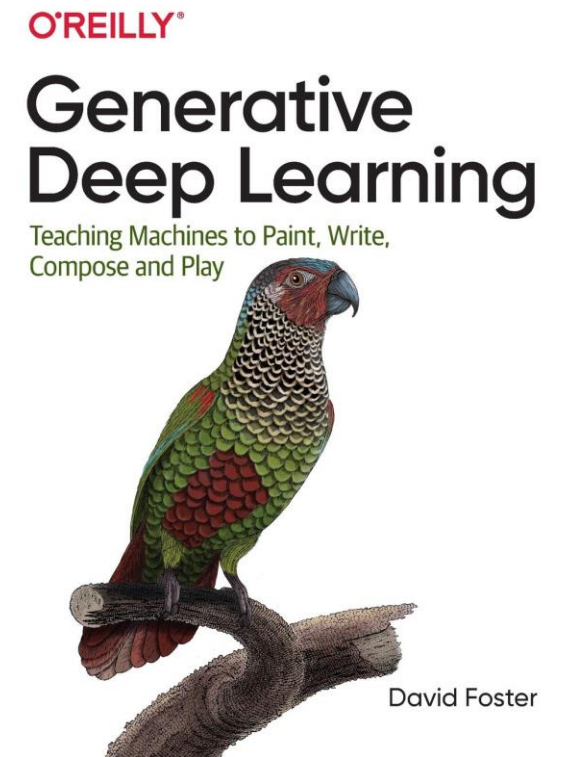
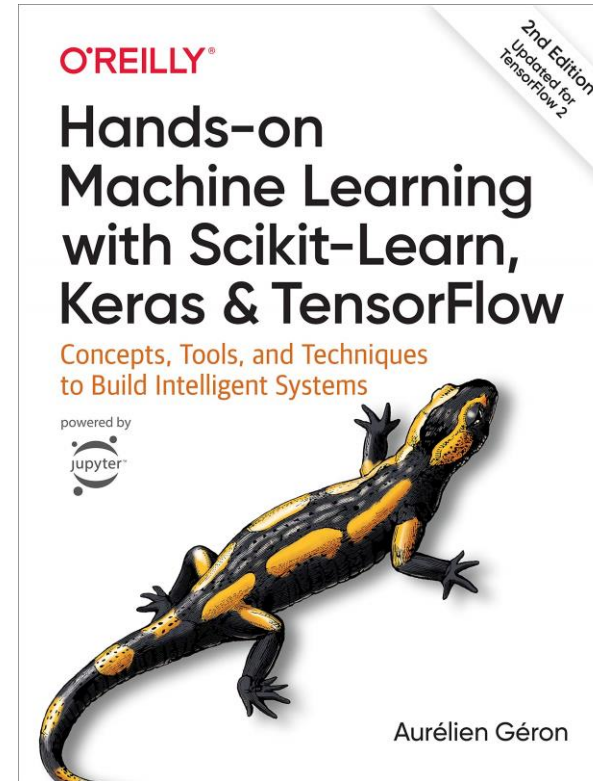
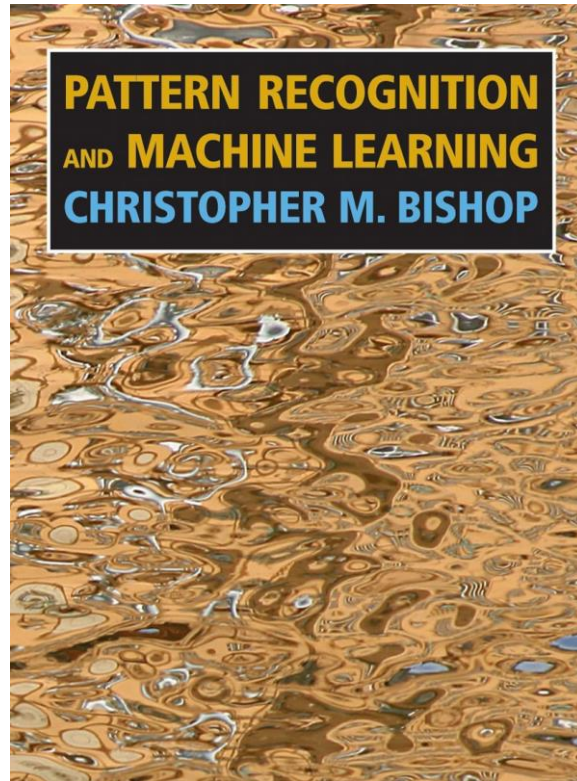
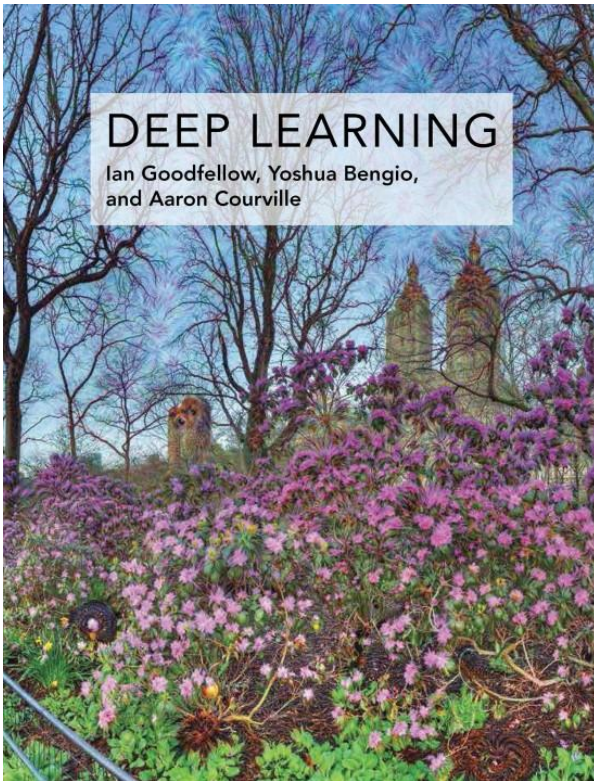


- Model deployment
- CI/CD pipelines
- Monitoring & triggering

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

Textbook

- No mandatory textbook for this course



Course Assessment

Component	Learning Outcomes	Weight
Participation	1, 2, 3, 4	5%
Assignments (2)	1, 2, 3, 4, 5	20%
Midterm	1, 2, 3, 4	30%
Final Project	1, 2, 3, 4, 5	45%

Participation (5%)

- Student participation will account for 5% of the final grade. How participation will be measured?
 - Students questions and answers during class
 - Students participation on the D2L discussion board
 - Students helping each other during class

Assignments (20%)

- Team-based – 5 people per team
- **Assignment 01 (10%):**
 - Proposing a garbage classification system based on images and natural language
 - **Due:** January 30th (midnight) | **Delivery method:** D2L dropbox
- **Assignment 02 (10%):**
 - Create a machine learning hands-on tutorial
 - **Due:** March 6th (midnight) | **Delivery method:** D2L dropbox

Assignment 1



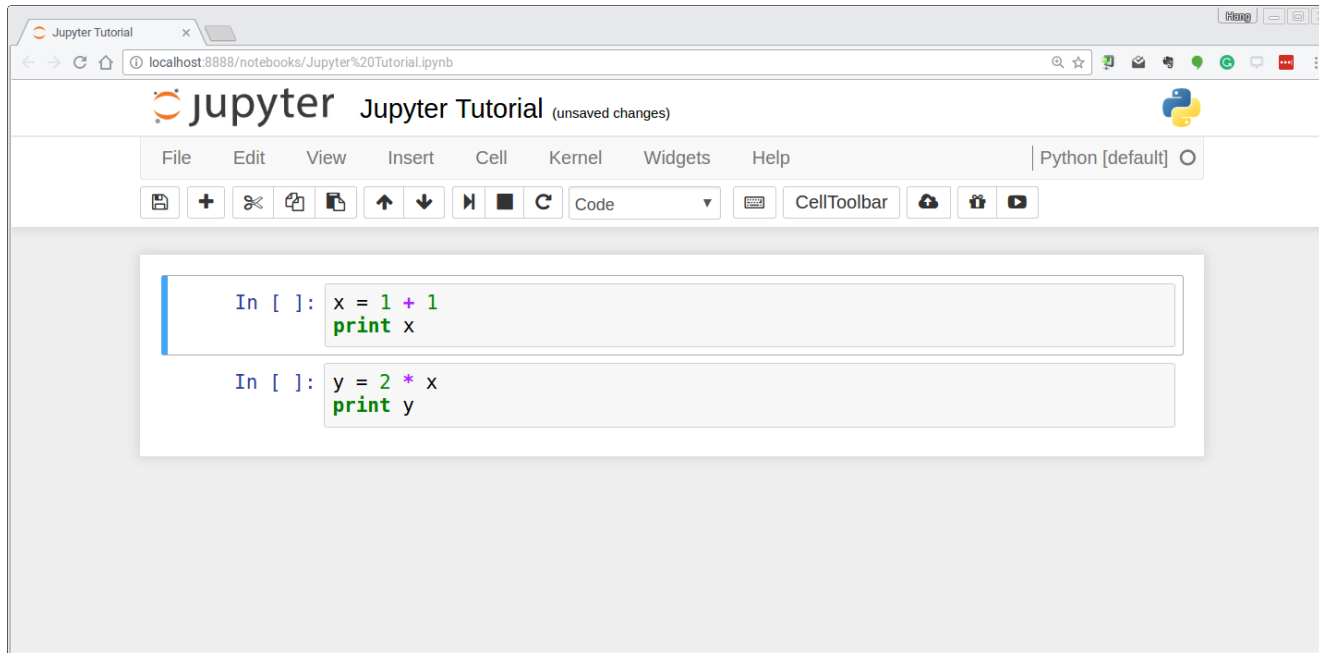
Clean cardbox -> blue trash bin



Greasy cardbox -> green trash bin

“Another potential issue is the class overlap for certain items. For example, a clean pizza box can be disposed of in the blue bin, however if it is greasy/covered in food toppings, it should be disposed of in the green bin. Depending on how the photo is taken, it may be impossible to know for sure which class this item belongs to.”

Assignment 02



- Develop a deep learning solution to the garbage classification problem used in assignment 1 (data will be provided)

Midterm (30%)

- Quizzes are individual
- Multiple choice
- A sample quiz will be provided for studying
- Content: all topics covered until the day of the midterm

- **Date:** February 28th in the classroom
- **Accommodation:** March 6th in the classroom (no class for those who already took the quiz)

Final Project (45%)

- 5-page report + 1 additional page only with references (if necessary)
- Report template
 - Overleaf - please make a copy for your team.
 - Microsoft word
- Final report due date: **March 27th at 9 am**
- Final project presentations: **March 27th to April 8th**
 - 5-minute presentation + 3 minutes for questions
 - Send slides one day before your presentation
 - Presentations are not graded but can help raise your final report grade

Grades

Letter Grade	Total Mark (T)
A+	$T \geq 95\%$
A	$90\% \leq T < 95\%$
A-	$85\% \leq T < 90\%$
B+	$80\% \leq T < 85\%$
B	$75\% \leq T < 80\%$
B-	$70\% \leq T < 75\%$
C+	$65\% \leq T < 70\%$
C	$60\% \leq T < 65\%$
C-	$55\% \leq T < 60\%$
D+	$50\% \leq T < 55\%$
D	$45\% \leq T < 50\%$
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>



UNIVERSITY OF CALGARY

Research Computing Services [Home Page](#)

The Programming Environment (Part 2)

- Python 3
- Python libraries:
 - NumPy
 - Matplotlib
 - Pandas
 - Scikit-learn
 - Tensorflow (version ≥ 2.0)
 - PyTorch
 - Weight and Bias
- Please have your programming environment in your computer or on Google Colab set up asap

Deep Learning Frameworks



**I hope you enjoy the
class 😊**

Questions?