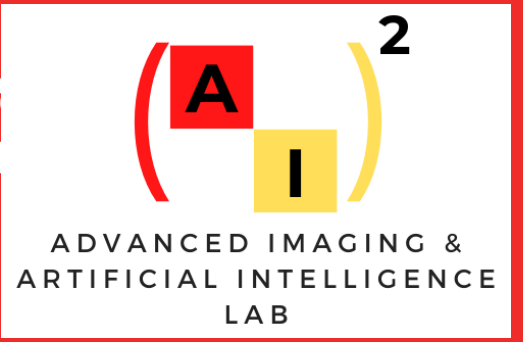
ENEL/ENEN 645 – Data Mining & Machine Learning



@lab_ai2

Overview of the course

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

W2025

UNIVERSITY OF CALGARY

Course Delivery

- Synchronous and in-person
  - MWF – 9:00 am to 9:50 am
  - Room ENE 241

- Use the D2L discussion board for questions
  - Alternative: Teams or slack channel?

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.

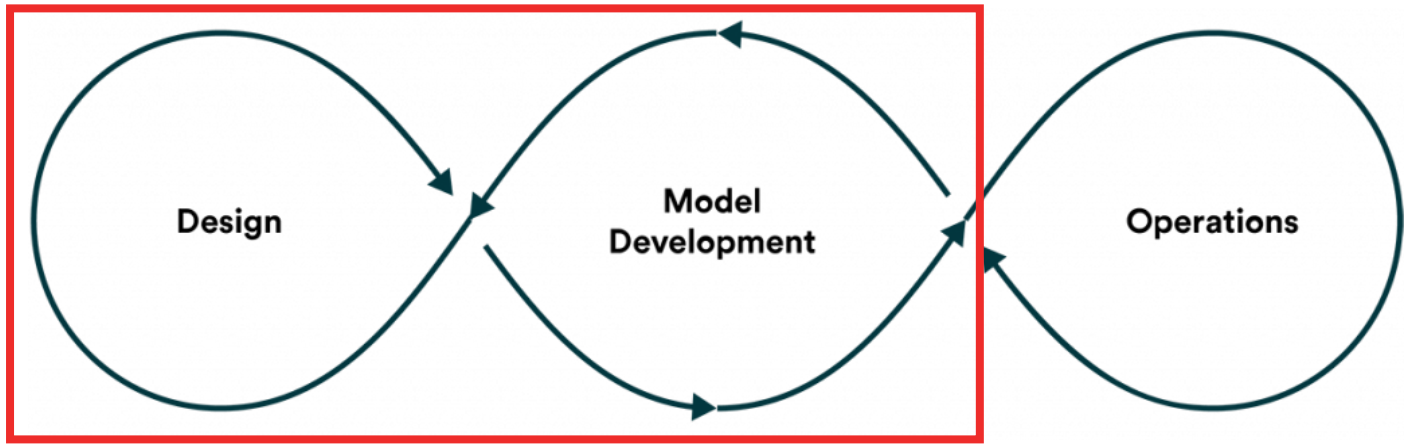
Course Syllabus (main topics)

- 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?



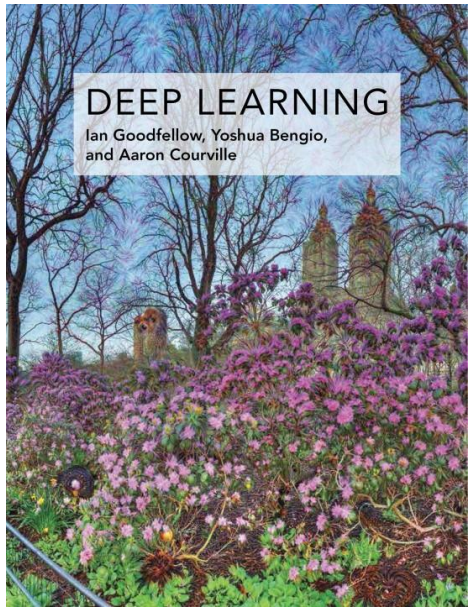
- Model deployment
- CI/CD pipelines
- Monitoring & triggering

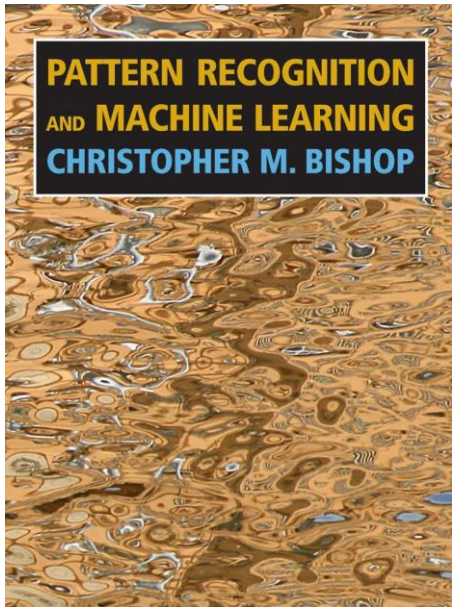
This course is about designing and developing machine learning models to achieve the best quantitative metrics* to the problems being modelled.

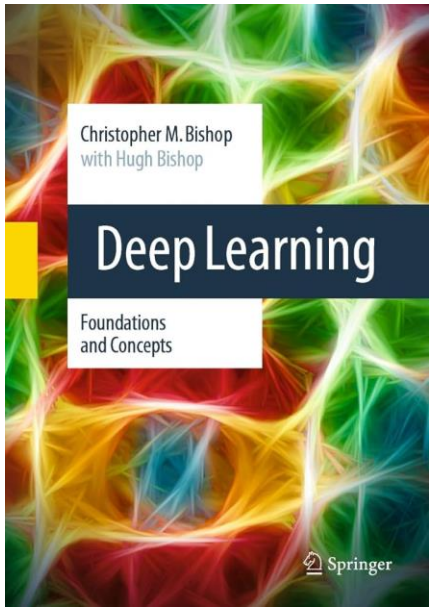
*Quantitative results are determined on a problem-by-problem basis.

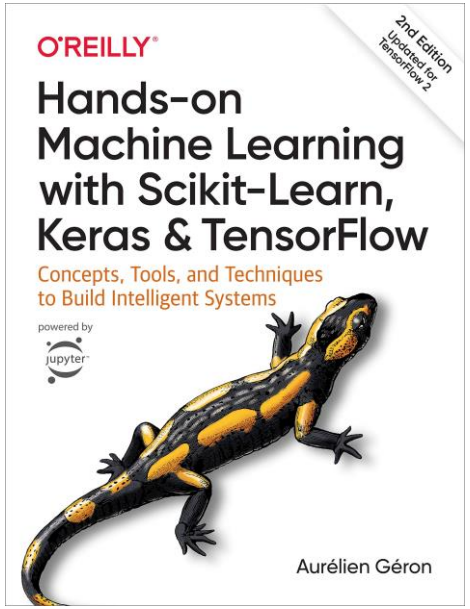
Textbook

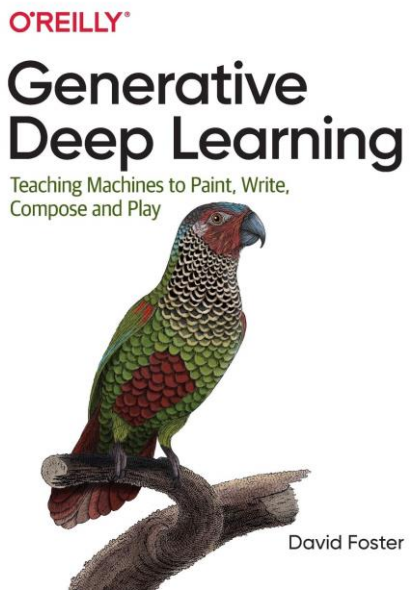
- No mandatory textbook for this course











Class GitHub repo: https://github.com/rmsouza01/ENEL645-F2024

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 will participation be measured?
  - Students helping with proposed activities during class
  - Students questions and answers during class
  - Students participation on the discussion channels
  - Students helping each other during class

Assignments (20%)

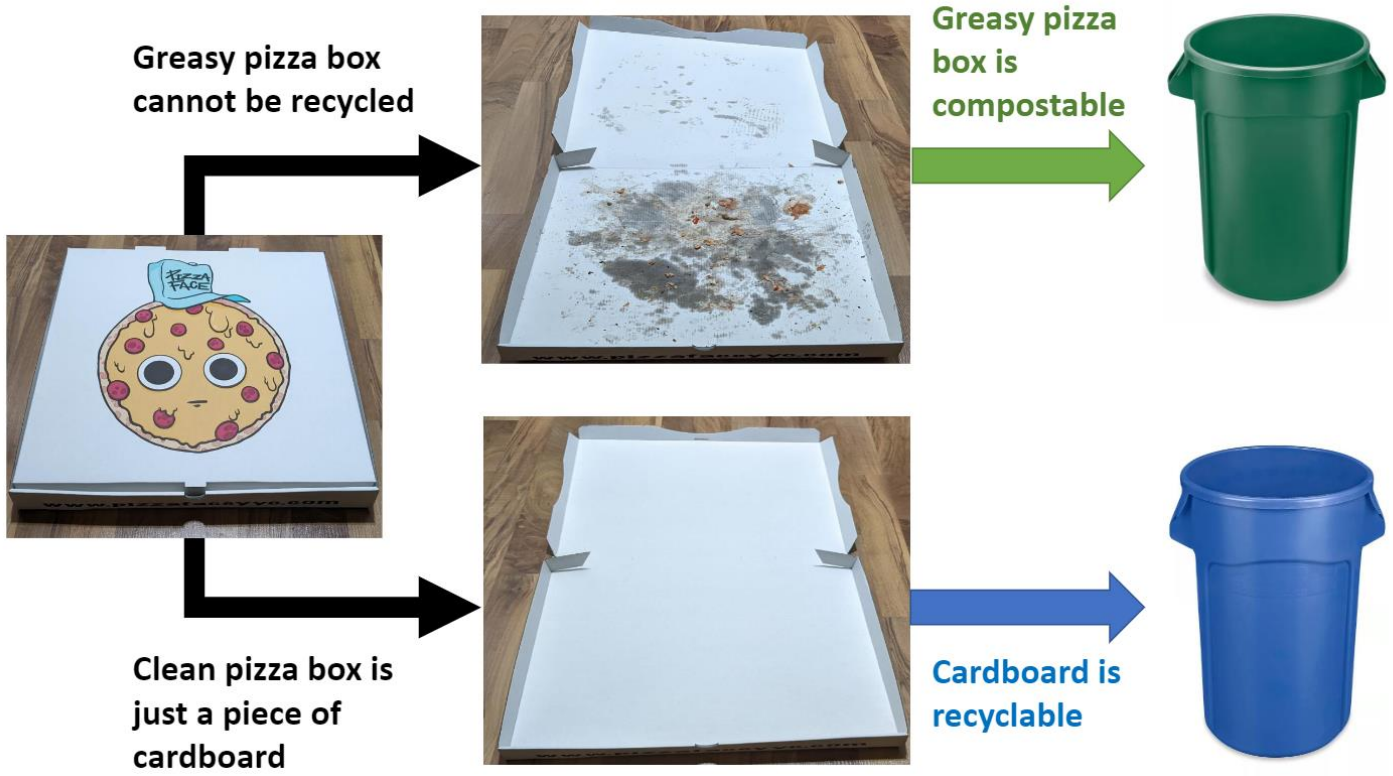
- Team-based – 4 people per team

- **Assignment 01** (10%)**:**
  - Proposing a garbage classification system based on images and natural language
  - **Due:** 31 January (midnight) | **Delivery method:** D2L dropbox

- **Assignment 02** (10%):
  - Propose a solution for the problem of assignment 01
  - **Due:** 14 February (midnight) | **Delivery method:** D2L dropbox

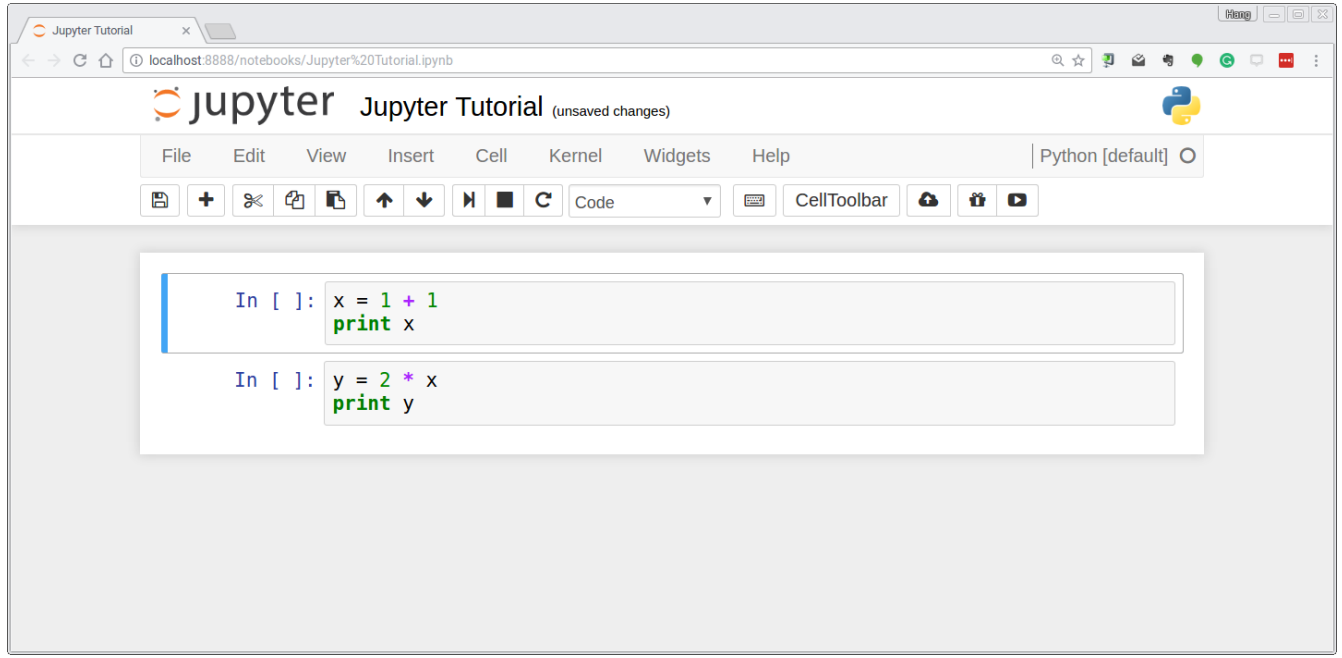
Assignment 1

- Design a garbage classification system based on images and text
- Use text to add context potentially not available in the image



UNIVERSITY OF CALGARY

Assignment 02



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

UNIVERSITY OF CALGARY

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:** 03 March 2025 in the classroom
- **Accommodation:** 10 March 2025 in the classroom

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: **04 April 2025 at midnight**

- Final project presentations: **TBD**
  - 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≥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/deep-learning

https://www.overleaf.com/project

UNIVERSITY OF CALGARY
Research Computing Services me_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?