

# Algorithmes d'apprentissage

non supervisé

01

## Introduction

# Au programme

---

- Tour de table
- Introduction à l'apprentissage non supervisé et ses principaux défis
- Organisation du cours et pédagogie
- Environnements logiciels du cours
- Littérature recommandée

# Tour de table

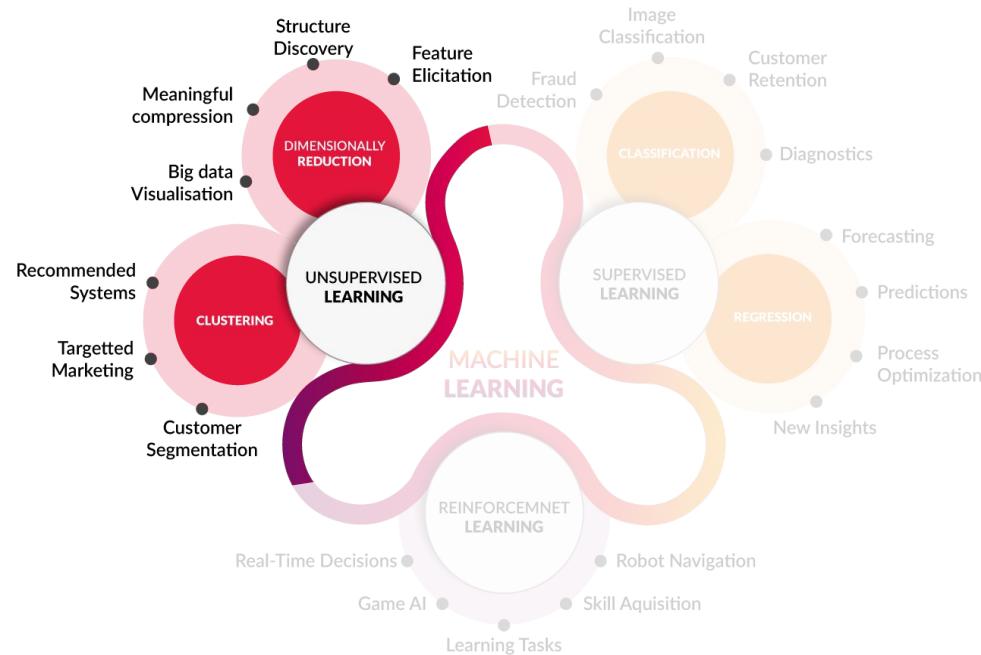
---

# **Introduction à l'apprentissage non supervisé et ses principaux défis**

---

# Introduction

---



Extrait du cours “Introduction à l’apprentissage automatique” de François Laviolette, CRDM

# Exemples d'applications (1/3)

Optimisation de portfolio



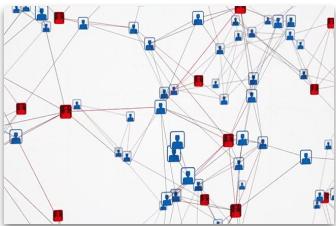
Centres de données



Astrophysique



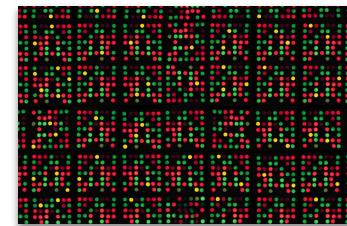
Réseaux sociaux



Marketing



Génétique



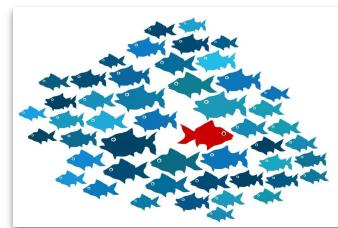
# Exemples d'applications (2/3)

---

Cybersécurité



Détection d'anomalies



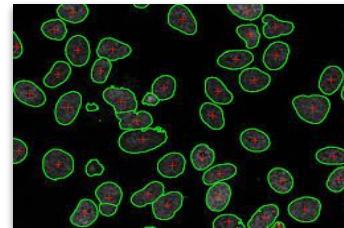
Commerce de détail



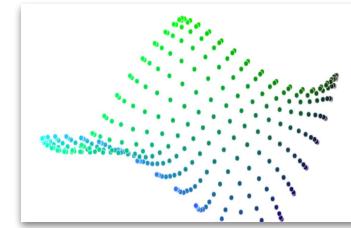
Recommandation



Vision numérique



Visualisation



# Exemples d'applications (3/3)

---

Télédétection



Ingénierie de données



Recherche de documents



IA Générative



LLMs



Et bien d'autres ...

# Le “gâteau” de Yann LeCun



Yann LeCun

Apprentissage  
non supervisé



## ■ “Pure” Reinforcement Learning (cherry)

- ▶ The machine predicts a scalar reward given once in a while.
- ▶ **A few bits for some samples**



## ■ Supervised Learning (icing)

- ▶ The machine predicts a category or a few numbers for each input
- ▶ Predicting human-supplied data
- ▶ **10→10,000 bits per sample**

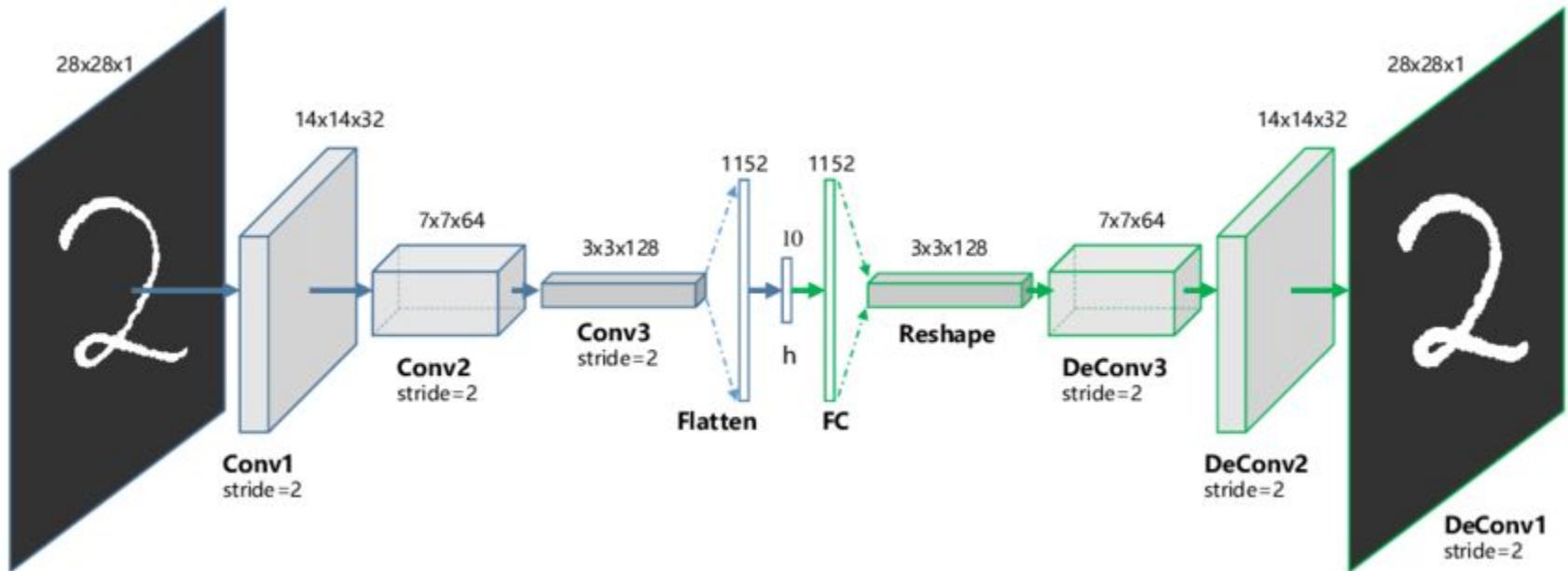
## ■ Unsupervised/Predictive Learning (cake)

- ▶ The machine predicts any part of its input for any observed part.
- ▶ Predicts future frames in videos
- ▶ **Millions of bits per sample**

■ (Yes, I know, this picture is slightly offensive to RL folks. But I'll make it up)

**L'apprentissage non supervisé attend toujours sa révolution, son “deep learning” ...**

# Deep learning “non supervisé”

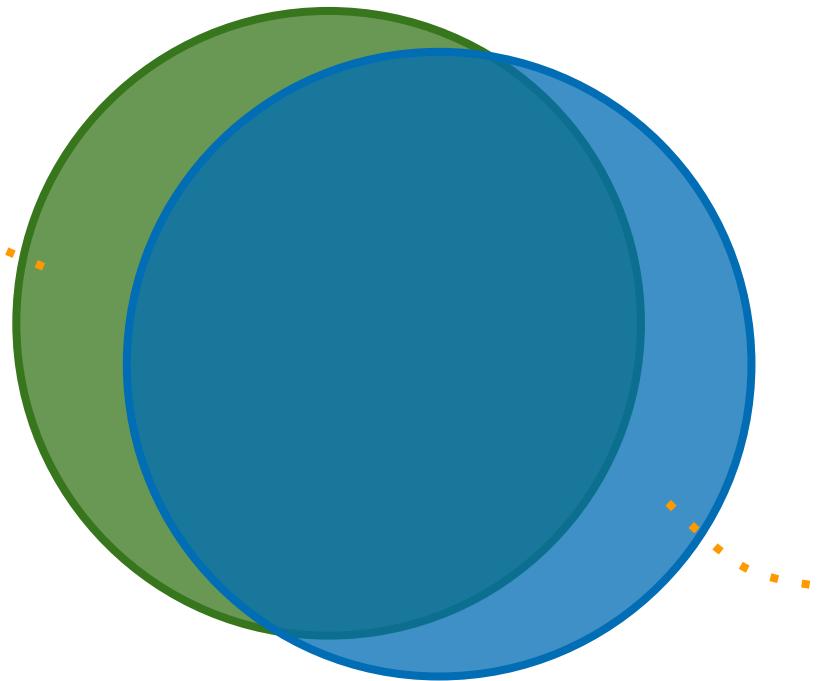


The structure of proposed Convolutional AutoEncoders (CAE) for MNIST

# Apprentissage non supervisé vs fouille de données

---

Apprentissage non  
supervisé



Fouille de  
données

# Pourquoi l'apprentissage non supervisé est “difficile” ?

- Comme son nom l'indique, le caractère non-supervisé signifie l'absence de variable dépendante ( $y$ )
- Les données sont donc non étiquetées ( $X$ )
- Il n'y a pas de bonne réponse à un problème et la subjectivité est souvent de mise
- Étant donné l'absence d'étiquettes, il est difficile, voir parfois impossible, d'évaluer la performance des algorithmes
- Les résultats doivent être interprétés en se basant sur des connaissances reliées au domaine d'application
- Mais tout ceci ne fait que rendre l'apprentissage non supervisé encore plus **passionnant !**

# Sujets non couverts par le cours

---

- **Réduction de dimension (ACP, t-SNE, ...)**
  - En partie couvert par 420-A55-SF
- **Détection d'anomalie**
  - En partie couvert par 420-A56-SF
- Les compétences correspondant à ces sujets (exemple: ACP) sont considérées acquises par l'étudiant et pourront être requises lors des examens à venir

# **Organisation du cours**

---

# Organisation du cours

---

- Il s'agit d'un cours de 60 heures
- Sauf mention contraire, les séances ont lieu sur Microsoft Teams
- Entre 30 minutes et 40 minutes de **pause** sont réparties sur la durée de chaque séance
- En dehors des cours, entre-aidez vous sur le salon de clavardage Teams. Restez polis et courtois ;-)
- Vos **commentaires** (positifs, constructifs) sur le cours sont les bienvenus !

# Pondération

---

**2 - 2 - 4**



# Évaluations

---

- Examen #1 (3.5 heures) - **30%**
- Examen #2 (3.5 heures) - **30%**
- Examen #3 (3.5 heures) - **30%**
- Évaluations formatives - **10%**

# Plan de cours

- Le plan de cours détaillé comprenant le calendrier est disponible sur **Lea**

Sommaire de la classe

**420-A58-SF gr. 12022**  
ALGORITHMES D'APPRENTISSAGE NON SUPERVISÉ  
mar 18:00, jeu 18:30

- Classe à distance** À distance
- Plan de cours** Distribué
- Documents et vidéos**
- Travaux et remises** Ajouter un énoncé
- Notes d'évaluation** Définir une grille
- Étudiants ayant utilisé Léa** 28 / 29 depuis les 10 derniers jours
- Absences et retards** Ajouter   
Consignation automatique Activer
- Forum de classe** Activer
- Communiqués** Ajouter
- Événements** mois à venir Ajouter



**FORMATION CONTINUE ET SERVICES AUX ENTREPRISES**

**Plan de cours**

**AEC de spécialisation technique en Intelligence artificielle (LEA.D1)**

Session : Hiver 2026  
Bloc : 3  
Cohorte : 4117  
Groupe : 12022

Aide pédagogique  
Individuelle (API) : Marie-Ève Gagné par MIO-Omnivoix  
Conseiller pédagogique : Eric Boudreault Courriel : [eboudreault@cfov.ca](mailto:eboudreault@cfov.ca)

Enseignant : Mikael Swawola Courriel : [mswawola@cfov.ca](mailto:mswawola@cfov.ca)

**ALGORITHMES D'APPRENTISSAGE NON SUPERVISÉ (420-A58-SF)**

Pondération : 2-4 Durée : 60 heures Unités : 2,67 Préalable(s) : aucun

**THÉMATIQUE GÉNÉRALE DU COURS**

Le cours Algorithmes d'apprentissage non supervisé (420-A58-SF), destiné aux étudiants du programme Spécialisation technique en intelligence artificielle (LEA.D1) comporte des cours magistraux, des démonstrations et des travaux pratiques pour acquérir les compétences de base permettant de comprendre les technologies nécessaires à la mise en place d'une solution d'apprentissage non supervisé.

De nombreux exercices et ateliers adressant les problématiques d'apprentissage non supervisé dans le cadre de domaines d'applications réelles permettent la mise en application des notions acquises. Le niveau de difficulté de ces travaux sera approprié aux objectifs du programme et les étudiants devront, au préalable, mettre en place un environnement adéquat.

**PLACE DU COURS DANS LE PROGRAMME**

Ce cours est le seul cours porteur de la compétence BBTG et est préalable aux cours Algorithmes d'apprentissage profond (420-A60-SF) et Préparation d'une solution d'IA pour la mise en production (420-A61-SF). Aucun cours n'est co requis.

Ce cours présente 60 heures d'enseignement, alternant séquences de théorie, démonstrations et travaux pratiques en classe virtuelle. L'étudiant devra s'attendre à fournir en moyenne 60 heures de travail personnel en dehors des heures d'enseignement.

PC\_420-A58-SF\_H2026\_4117\_MS

1

# Structure du cours

---

- Le cours est divisé en 3 parties
  1. Partitionnement de données
  2. Fouille de données
  3. Systèmes de recommandation

# Pédagogie

---

- Les ateliers seront réalisés et corrigés en classe.
- Les séances seront toujours enregistrées.
- Les **supports de cours** seront postés sur le dépôt GitHub du cours au début de chaque séance

- Le mode d'enseignement est **synchrone**, c'est-à-dire:
  - Sauf exception, la présence aux séances est **obligatoire**
  - Les séances sont enregistrées, mais ne devrait servir qu'à s'y référer pour consolider les connaissances acquises après avoir assisté aux différentes séances
  - Il est important de ne pas raisonner dans l'autre sens, à savoir remettre à plus tard l'écoute d'un cours sachant que celui-ci est enregistré. Ce n'est pas un MOOC ! De plus, rattraper un cours de 3h nécessite généralement beaucoup plus que 3 heures !

# **Environnements logiciels du cours**

---

# Environnements

---

- L'environnement du cours comporte Linux (WSL2), Docker Desktop et VSCode
- Jupyter Notebook et le langage de programmation Python (3.12.3) seront utilisés
- **À ce stade de la formation, il est important que vous soyez en mesure d'installer et configurer ces environnements de manière autonome.**

# Dépôt du cours

---

- Tous les exercices et jeux de données seront accessibles au fur et à mesure sur le **dépôt GitHub du cours:**  
<https://github.com/mswawola-cegep/420-a58-sf-gr-12022-hiver-2026.git>
- Utilisez l'image Docker pour assurer votre environnement d'une complète compatibilité avec les ateliers du cours

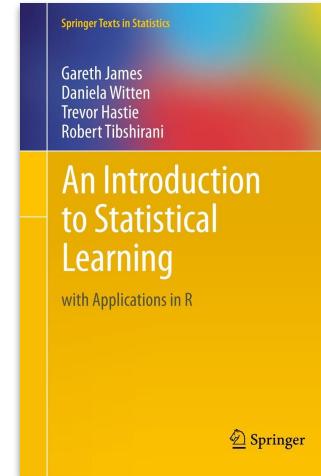
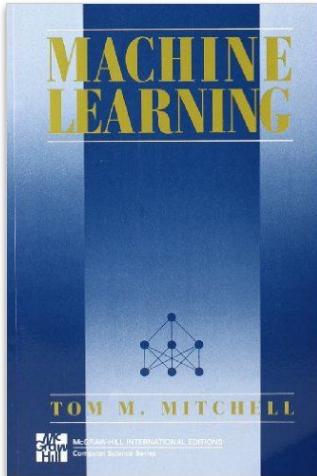
## Littérature recommandée

---

# Littérature recommandée

---

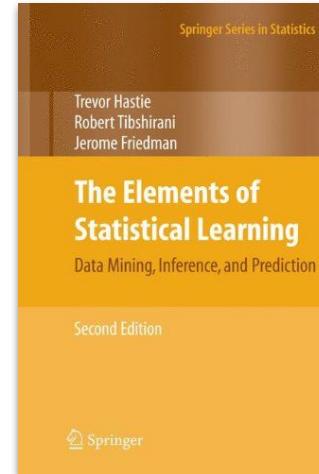
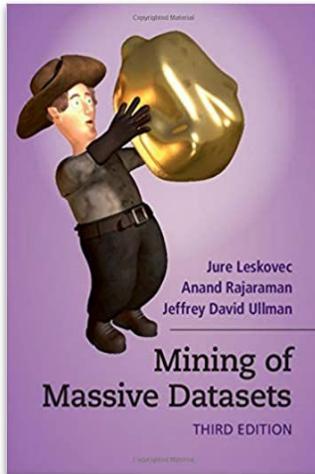
- **Machine Learning**, Tom Mitchell  
McGraw-Hill Higher Education
- **An Introduction to Statistical Learning with Applications in R**, G. James, D. Witten, T. Hastie, R. Tibshirani



# Littérature recommandée

---

- **Mining of Massive Datasets**, J. Leskovec, A. Rajaraman, J.D. Ullman
- **Elements of Statistical Learning Second Edition**, T. Hastie, R. Tibshirani, J. Friedman



# Littérature recommandée

---

- **Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems**, Aurélien Géron
- **Advances in Financial Machine Learning**, Marcos Lopez De Prado

