

M2 Data Science – M2 Intelligence Artificielle

Data Mining

Introduction



Présentation rapide

- Antoine Richard
 - Chercheur en IA appliquée au domaine de la Santé
 - Centre pour l'Innovation en Cancérologie de Lyon (CICLY) – HCL / UCBL
- Principaux sujets de recherches:
 - AI/ML pour la Santé
 - Protection de la vie privée
 - Ethique des algorithmes
- Liens utiles
 - Site web: <http://a-t-richard.github.io> (<-- tous les supports de cours seront ici)
 - Mail: antoine.richard@chu-lyon.fr

Concernant ce cours

- Repris du cours de Rémy Cazabet: <https://cazabetremy.fr/Teaching/DSIA/DM.html>
- Apprentissage Par Problème/Projet (APP)
 - Groupes de 5 (tirés au sort), avec au minimum:
 - Un·e maître du temps
 - S'assure que personne ne monopolise la parole et que tout le monde participe équitablement
 - Un·e scribe
 - S'assure de synthétiser le résultat de la discussion et que tout le monde est en accord avec la retranscription
 - Chaque session de nouvelles données à analyser en équipe
 - Introduction de nouveaux concepts au fur et à mesure
- Supports de cours disponible ici:
 - <http://a-t-richard.github.io/teaching/data-mining>

Concernant l'usage de l'IA Générative

« Le cerveau, c'est comme un muscle, il a besoin de s'exercer pour s'améliorer.
Et utiliser ChatGPT pour apprendre, c'est comme faire des tractions avec un transpalette »

Donc, l'usage de l'IA Générative est déconseillé.

Mais si vous souhaitez tout de même l'utiliser:

- **Soyez honnête et transparent:**
 - Dites quand, pourquoi et comment vous l'utilisez
- **Soyez critique:**
 - Vérifiez et questionnez les résultats

Problème de définition



Pas de définition claire

Objectif: trouver une définition

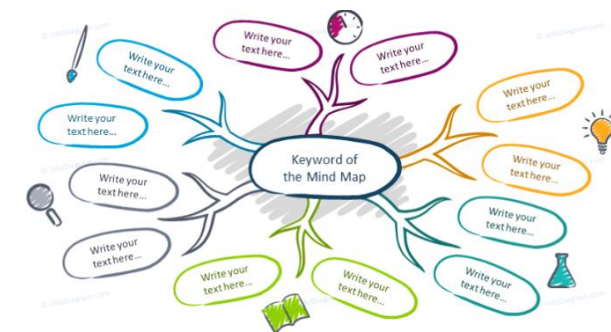
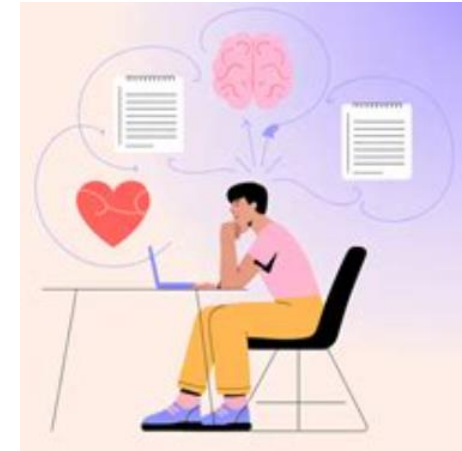
- C'est quoi « **une donnée** » ?
- Qu'est-ce que ça signifie « **miner** » ?
- Qu'est-ce que c'est le « **minage de données** » ?

Favoriser l'intelligence collective

Explorer et exploiter le champs des idées

Méthode de réflexions intra / inter personnels:

1. Réfléchissez chacun.e de votre côté, sans interactions, et notez vos idées
2. Au bout de 5-10 minutes, mettez en commun vos idées
 - Ex: faites une *mindmap* commune
3. Débattez des zones d'intérêt à « creuser » un peu plus
4. Recommencer à (1.) jusqu'à ce qu'il n'y ait plus régulièrement de nouvelles idées qui émergent



1. [Hills et al. \(2015\) – « Exploration versus exploitation in space, mind, and society » in Trends in Cognitive Sciences](#)
2. [Bernstein, Shore and Lazer \(2018\) – « How intermittent breaks in interaction improve collective intelligence » in Proceedings of the National Academy of Sciences](#)

Favoriser l'intelligence collective

Arriver à un consensus

Méthode « Delphi » simplifiée:

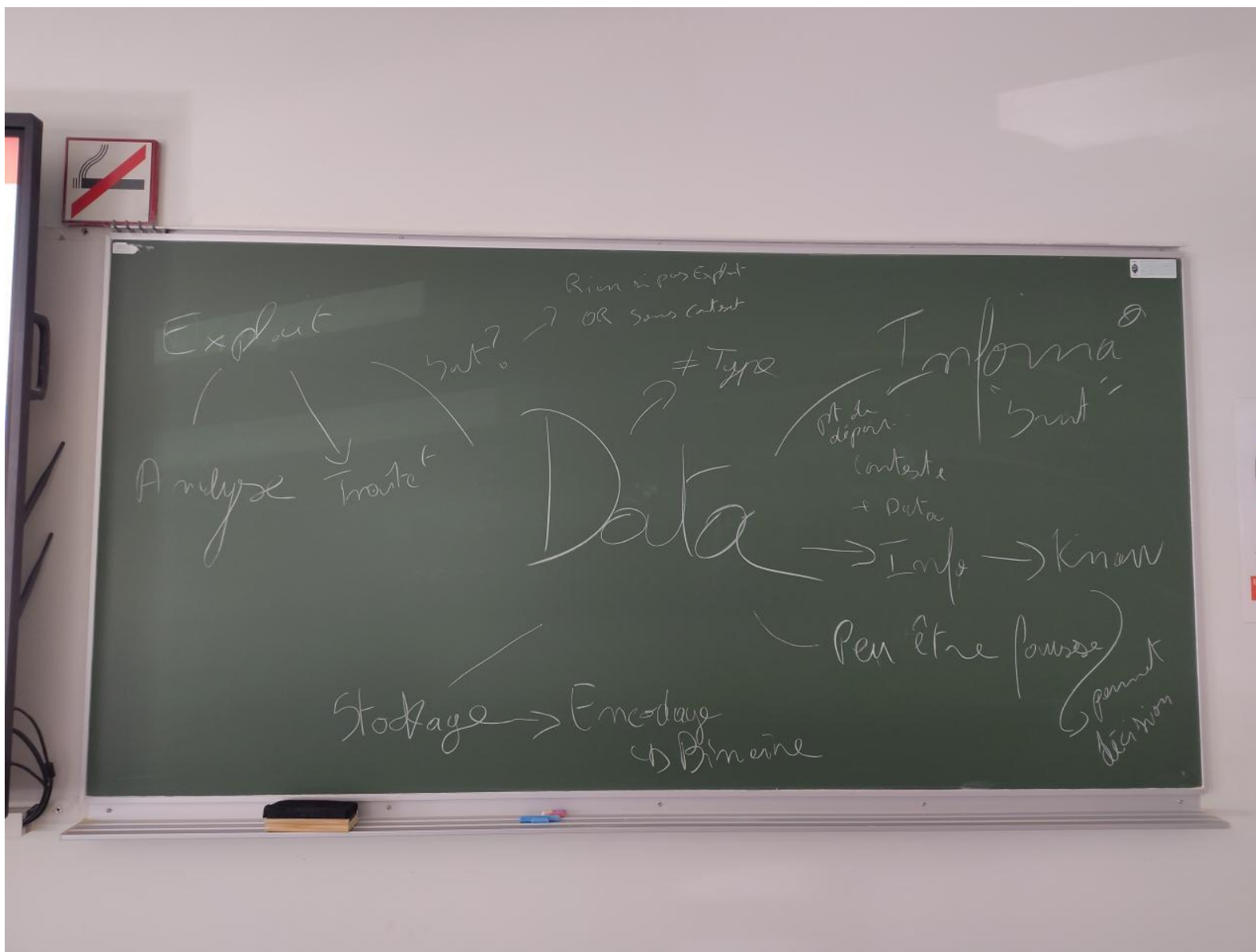
1. Débattez en vous assurant que le temps de parole est réparti équitablement
 - Pro-tips:
 - Évitez que se soit toujours la même personne qui parle en premier
 - Essayez de vous mettre à la place des autres, de comprendre leur point de vue et de considérer le problème sous différents angles
2. Synthétisez une proposition (ici: la définition)
3. Votez si vous êtes en accord ou non avec la proposition (idéalement anonyme)
 1. Si accord unanime → La proposition est validée
 2. Sinon, si il reste une ou deux personnes pas en accord mais ok pour valider la proposition → La proposition est validée
 3. Sinon, retournez à (1.)

1. [Al-Qutaish \(2025\) – « Delphi Method: a comprehensive literature review » in *International Journal of Software Engineering*](#)
2. [Wolley et al. \(2010\) – « Evidence for a Collective Intelligence Factor in the Performance of Human Groups » in *Science*](#)

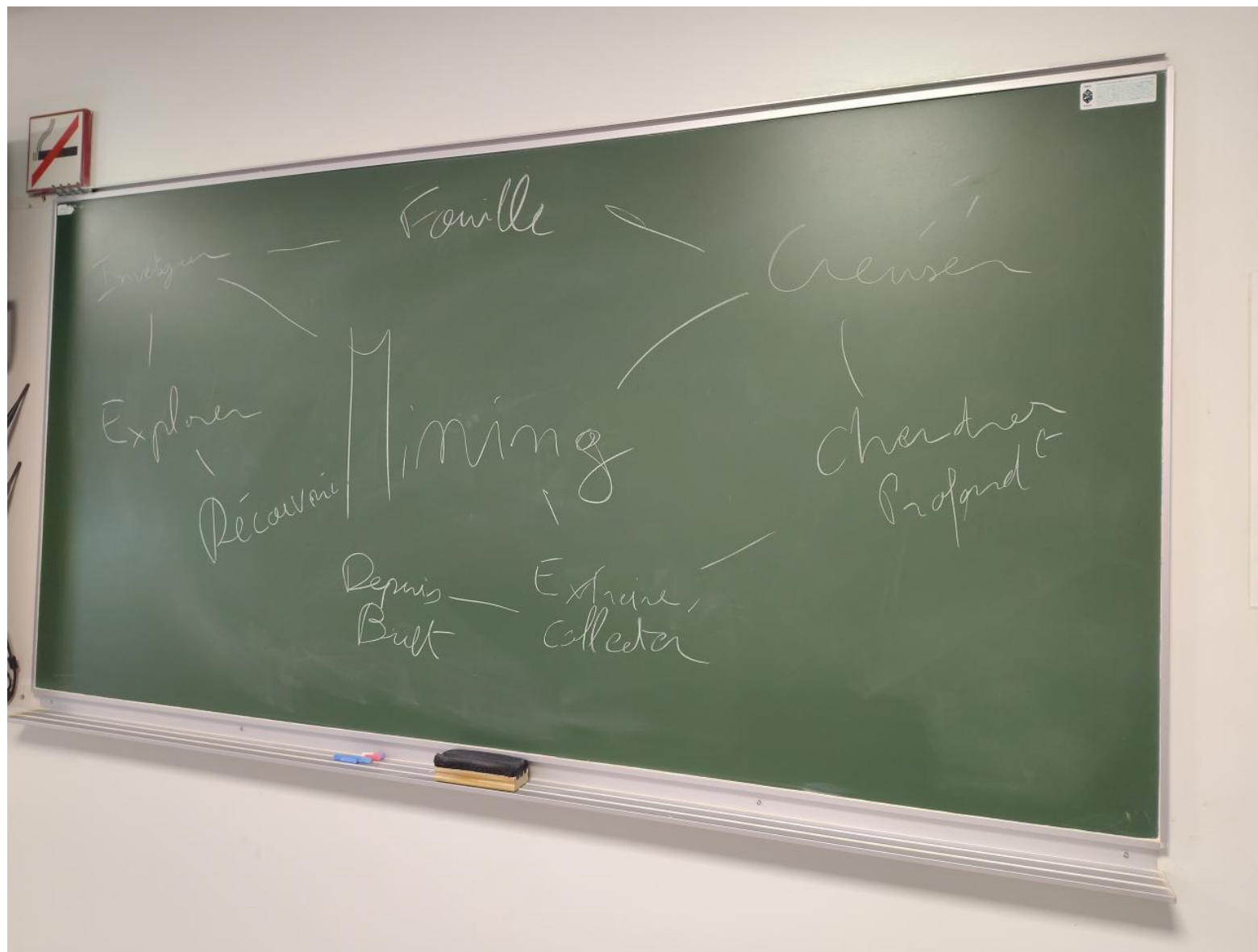
Résultats - Définition

- « **une donnée** » c'est: « un élément collecté dans le but d'essayer de stocker une information relatant une observation, peut prendre plusieurs formats/type, qui peut être exploité (analyse, traitement), avec un cycle de vie (collecte -> exploitation -> suppression). »
- « **miner** » c'est: « Investiguer/explorer/fouiller profondément un ensemble brut pour y découvrir et en extraire quelque chose »
- Le « **minage de données** » c'est : « Extraction informatique, depuis des ensembles massifs de données brutes (BigData), d'informations afin d'en tirer des connaissances via l'analyse, pour prendre des décision »

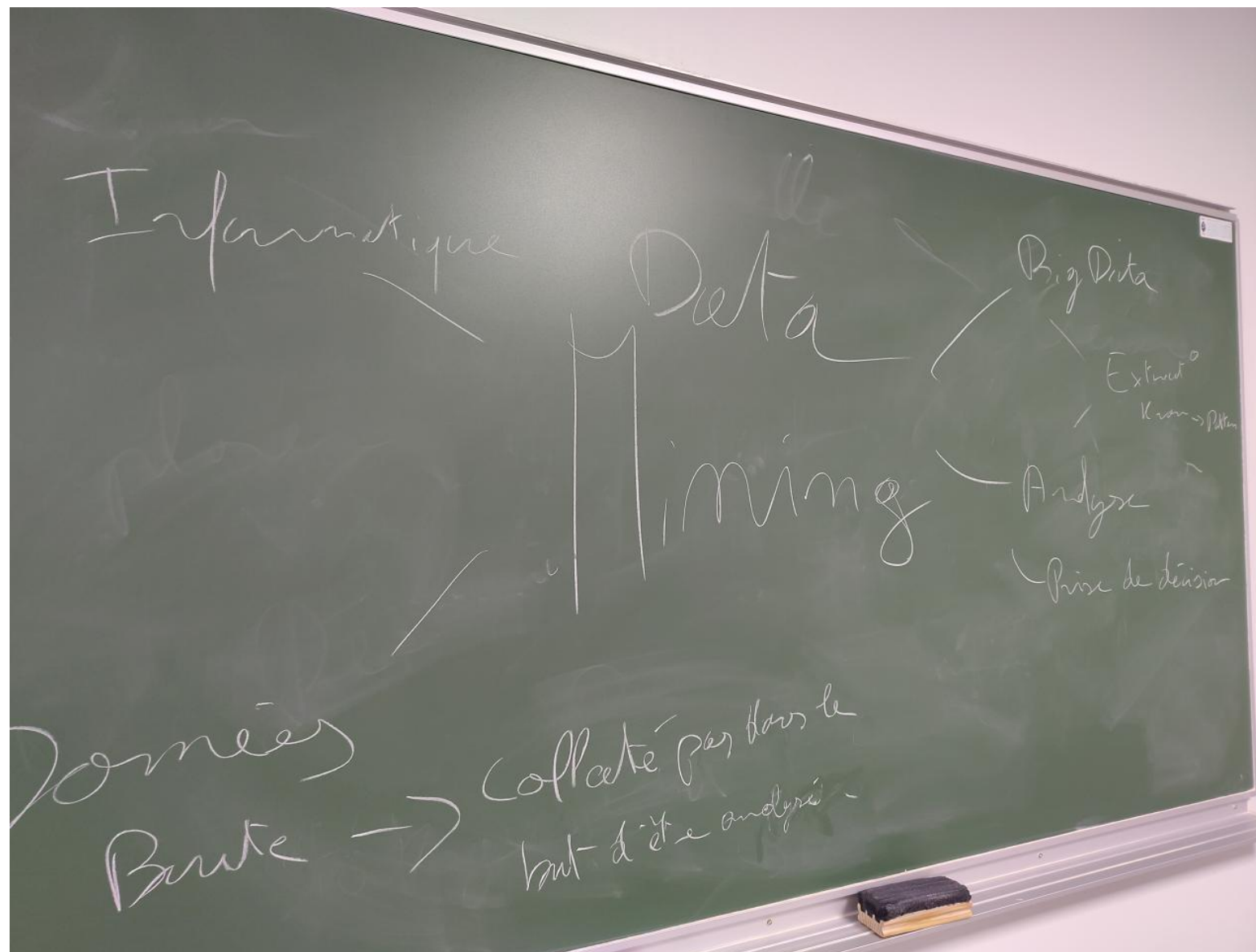
Résultats – Mindmaps - Data



Résultats – Mindmaps - Mining



Résultats – Mindmaps – Data Mining



À suivre

- Cours:
 - Cours 1 (14/09): Les bases de la fouille de données
 - Cours 2 (15/09): Ordonner et regrouper des données
 - Cours 3 (05/10): Identifier des patterns et des anomalies
 - Cours 4 (06/10): Miner des graphes et des processus
 - Cours 5 (12/10): Miner des données spatiales et temporelles
 - Cours 6 (13/10): Miner des données textuelles
- Examen (03/11):
 - 50 % de la note finale
 - 12 pts (60%) sur des exercices vu en cours, 8pts (40%) sur un problème inconnu
- Projet :
 - Travail en groupe
 - Un ensemble de jeux de données à fouiller
 - Écrire un rapport sous la forme d'un notebook Jupyter (Python)
 - 50% de la note finale