

Resumé Data mining:

DM: est une processus utilisé pour extraire de données.
l'objectif du DM: est de prendre des décisions basés sur ces données.

les Types de DM:

- Classification: identification des caractéristiques des objets dans les données.
- Association: Découverte de relation entre différents éléments de données.
- ~~Clustering~~ clustering: Regroupement de données en fonction de leur similitude.
- Analyse d'évolution et de déviation: suivi des changements dans les données ~~en fonction de leur similitude~~ au fin du Temps.

* Méthode de la Technique de DM pour l'exploration de données:

- Classification
- Clustering
- Association

• les Relations:

DM & Big DATA
Big DATA: fournit des données et les outils, Tandis que DATA mining utilise des techniques pour en extraire de connaissances.

DM & Analyse de données
Analyse des méthodes statistiques pour comprendre de données, Tandis que le DM utilise des modèles cachés à l'aide d'algorithmes avancés.

1

DM & Intelligent artificiel.

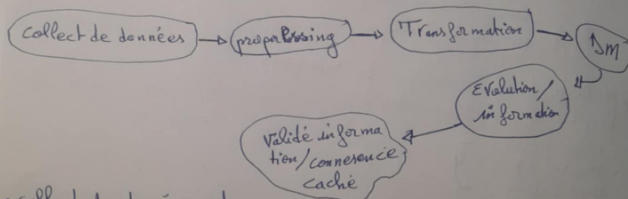
le DM utilise des techniques de DM, comme l'apprentissage automatique pour entraîner de modèles cachés et améliorer l'analyse de données.

DM & IR.

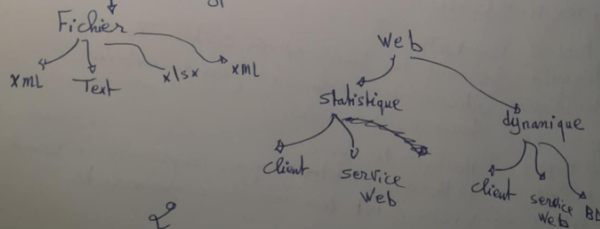
DM: ne recherche pas des informations mais se recherche des patterns cachés.

- KDD (Knowledge data discovery) est un processus:
- * pré-traitement (pre-processing)
- * DATA mining
- * post-traitement (post-processing).

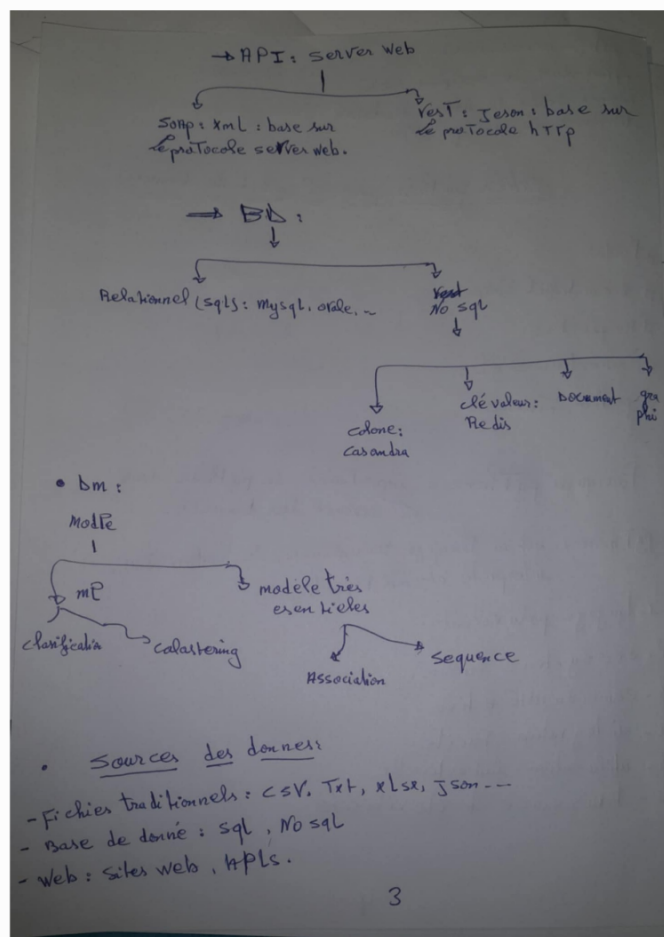
PIPELINE de DM:



• collect de données: type



2



- outils nécessaires:

- Framework de scraping web.
- connecteurs de bases de données.
- librairies de manipulation

Outils python pour la collect de données:

- Pandas:
- BeautifulSoup soup:
- Requests:
- connecteur sql:
- python Finance

Pourquoi python: importance de python dans la science des données.

Python: est un langage très avancé, de haut niveau, interprété et multi paradigme.

- 1- Langage polyvalent:
- 2- Ecosystème riche.
- 3- communauté active.
- 4- intégration facile
- 5- utilisation industrielle
- 6- puissance et flexibilité.

4

Règle d'association : Analyse de paniers d'achat
Modèle de commerce.

Donné : est un ensemble de transaction et
chaque transaction contient de élément

Données

T	Article	élément	T	Support (satur)
1	Pain, lait	pain	1, 4, 5	$\frac{3}{5}$ 60%
2		lait	1, 2, 4, 5	$\frac{4}{5}$ 80%
3	Pain, beurre, oeufs, jus	Beurre	2, 3, 4	$\frac{2}{5}$ 40%
4	lait, jus, thé, beurre	Jus	2, 3, 4, 5	$\frac{4}{5}$ 80%
5	pain, lait, jus, beurre	thé	3, 5	$\frac{2}{5}$ 40%
	pain, lait, jus, thé	oeufs	2	$\frac{1}{5}$ 20%

$$\text{Support} = \frac{\text{nbr de transaction de } x}{\text{Total } T}$$

$$\text{Confiance} = \frac{\text{Support}(x \cup y)}{\text{Support } x}$$

$$\begin{aligned} & (\text{pain, lait}) \rightarrow 1, 4, 5 \rightarrow \frac{3}{5} = 60\% \\ & (\text{pain, lait, thé}) \rightarrow 1, 4, 5 \rightarrow \frac{2}{5} = 40\% \end{aligned}$$

$$\{ \text{pain} \} \rightarrow \{ \text{lait} \}$$

$$C = \frac{60\%}{80\%} = \frac{3}{4} = 75\%$$

$$\rightarrow \{ \text{pain, lait} \} \rightarrow \{ \text{thé} \}$$

$$S = \frac{1}{5} = 20\%$$

$$S = \frac{3}{5} = 60\%$$

$$C = \frac{0,2}{0,6} = \frac{1}{3} = 33,3\%$$

$$\rightarrow \{ \text{pain, lait, jus} \} \rightarrow \{ \text{thé} \}$$

$$S = \frac{1}{5} = 0,2 = 20\%$$

$$S_x = \frac{2}{5} = 40\%$$

$$C = \frac{20\%}{40\%} = \frac{1}{2} = 50\%$$

5

- groupement élément \rightarrow antécédent.

- élément approuvé \rightarrow support.

Lift:

si Lift est supérieur à 1 ≥ 1 on a \geq sur
et sinon cette règle est permise si est
 ≥ 1 peut être au hasard.

Support: par rapport à l'ensemble de
transaction

Confiance: par rapport à l'élément
présent.

6

Lift=confiance/support

A_b

