

1) l'exploration de données
sur quel type de données ?

Pourquoi Python dans data mining et ML ?

Comment on peut récupérer les données ?
Collecte \rightarrow Traitement \rightarrow Transformation \rightarrow data mining \rightarrow Prendre décision
Collecte \rightarrow Prétraitement (Data cleaning) \rightarrow Transformation (AR-2 TRS) \rightarrow data mining \rightarrow Prendre décision (interprétation)
(importantes données) (qualitatif)

- classification : Catégoriser des données
- régression : Prédire des valeurs
- clustering ou regroupement : grouper sans catégories

Exploration de données : Analyse, découvrir, extraire

sur quel type de données : données structurées et non structurées
(vente, données clients), non structurées (Emails, Vidéo, photos)

Donnée semi-structurée (XML, JSON)

Base de données : décisionnelle enlève des données dans
BD transactionnel grâce à la vision

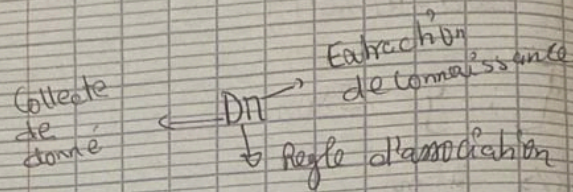
BD avancée : ceux qui gèrent les données complexes non structurées

Pourquoi Python

- Facile à apprendre : Python est simple d'utilisation
- Riches bibliothèques : Pandas, NumPy
- Analyse rapide possible : Gestion de grandes données
- supporte IA facilement : utilise pour machine learning
- Communauté très active : beaucoup d'aide disponible

data csv

year	Marke	Model
1987	Ford	E310
2000	Mercury	Cougar



Analyse de paniers d'achat

Articles

- 1 pain, lait
- 2 pain, beurre, oeufs, jus
- 3 lait, jus, beurre, thé
- 4 pain, lait, jus, beurre
- 5 pain, lait, jus, thé

Règle d'association

X → Antécédent

Y → Conséquence

trois métrique

- support: permet de mesurer la fréquence d'un

confiance,

- lift

support - soutien = $\frac{\text{nombre de transaction de } x}{T}$

sous ensemble
ou
élément

élément	support	\bar{I}	
Pain	4/5	1, 2, 3, 4	80%
Lait	4/5	1, 2, 3, 4	80%
Beurre	3/5	2, 3, 4	60%
Jus	4/5	2, 3, 4, 5	80%
Thé	2/5	3, 5	40%
Oufs	1/5	1	20%
Pain, lait	3/5	1, 4, 5	60%
Pain, lait, thé	1/5	1	20%
Pain, lait, jus, thé	2/5	4, 5	40%

$$\text{Confian}(x \rightarrow y) = \frac{\text{support}(x \cup y)}{\text{support}(x)}$$

$$\left\{ \begin{array}{l} \text{Pain, lait } y \\ S = 0,6 \end{array} \right\} \rightarrow \left\{ \begin{array}{l} \text{Thé } y \\ S = 0,4 \end{array} \right.$$

$$\text{pain, lait, thé} = 0,2$$

$$C = \frac{S(P, L, T)}{S(P, L)} = \frac{0,2}{0,6} = \frac{1}{3} = 33,33$$

st Pe
 une créance
 ou variable.
 tenant ou amené à
 positif à taux fixe
 à ce type de risque
 théorique de
 du taux d'intérêt

$$\begin{aligned}
 \{ \text{pain, lait, jus} \} &\Rightarrow \{ \text{thé} \} \\
 S = \{ P, L, J \} &= 0.4 \\
 S = \{ P, L, J, T \} &= 0.2
 \end{aligned}$$

$$\text{Confiance} = \frac{0.2}{0.4} = 50\%$$