

résumé data mining

1^{er} → c'est quoi data mining ?

→ data mining : est le processus d'extraction de modèle et de connaissance des ensembles de données ~~des~~ cachés.

→ data mining s'applique à différents types de données :

- 1 → Base de données relationnelles,
- 2 - Entre pôles de données.
- 3 - Base de données transactionnelles.
- 4 - Système de base de données avancées.

→ quels types de modèles peuvent être extraits ?

- 1 - classification
- 2 - Analyse d'association
- 3 - classification et prédiction
- 4 - Analyse de clustering
- 5 - Analyse de évolution et de déviation

2^{er} → pourquoi python important dans data mining ?

- 1 - langage polyvalent.
- 2 - Écosystème Riche.
- 3 - intégration facile.
- 4 - puissance et flexibilité.

3° → l'ensemble du pipeline d'exploration de données

collecte de données → preprocessing → transformation → interprétation
→ data mining → interprétation



Outils python pour la collecte de données

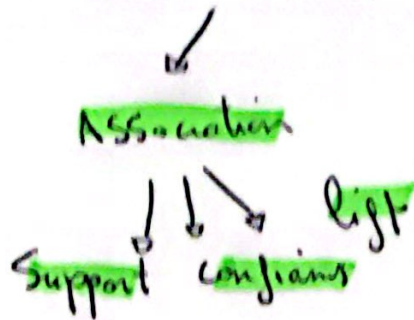
- pandas : lecture de fichiers CSV, excel, SQL.
- BeautifulSoup : Analyse HTML et XML
- requests : récupération de données via APIs et service web
- connecteurs SQL : mysql, python, sqllite3
- yahoo finance.

→ utilisation pandas pour lire un fichier :

```
import pandas as pd  
df = pd.read_csv('nom_fichier.csv')  
print(df.head(1))
```


Regles d'association

data mining



$X \longrightarrow Y$
Antecedent consequence

1- Support \rightarrow frequency = $\frac{\text{Nombre transaction } X}{\text{total transaction}}$

2- confiances $\& X \rightarrow Y$ = $\frac{\text{Support } \{X \cup Y\}}{\text{support } (X)}$

3- $\text{lift } (A \rightarrow B) = \frac{\text{confiance}}{\text{support}}$

$\rightarrow \text{lift} > 1$: Sont positivement associes.

$\rightarrow \text{lift} = 1$: independants.

$\rightarrow \text{lift} < 1$: negativement.

Data Mining :

Q₁ : C'est quoi Data Mining ?

R₁ : Data Mining est une recherche d'information avancée

Q₂ : C'est quoi Big Data ?

R₂ : Big data c'est une grande quantité de données complexes qui nécessitent des outils spéciaux pour être stockées, traitées et analysées.

Q₃ : Relation entre Data Mining et Big Data ?

R₃ : ~~Big Data se concentre sur le stockage~~

R₃ : Big Data fournit les données et les outils, tandis que data mining utilise des techniques pour en extraire des connaissances.

Q₄ : Relation entre Data Mining et Analyse des données

R₄ : L'analyse des données utilise des méthodes statistiques pour comprendre les données, tandis que le data mining extrait des modèles cachés à l'aide d'algorithmes avancés

Q₅ : Relation entre DM et Intelligence artificielle ?

R₅ : Le DM utilise des techniques d'IA, comme l'apprentissage automatique pour extraire des modèles cachés et améliorer l'analyse des données

Q₆ : Méthodes et techniques de DM pour l'exploration des données ?

R₆ : Classification, clustering, et association & régression

Q₇ : Pourquoi python important dans le data mining ?

R₇ : Le python c'est un langage très avancé, de haut niveau et interprété, ce qui le rend adapté à l'analyse de données ou et ou data mining.

Q₈ : KDD