

NOM :Ahmed salem /Mohamed

MATRICULE :22273

RÉSUMÉ SUR DATA MINING

*Qu'est-ce que le DATA MINING

*L'Objectif du DATA MINING

*Types des données

*L'importance de python en DATA MINING

*Pipeline d'exploration de données

*Règle d'association

Qu'est-ce que le DATA MINING :

Le Data Mining (ou exploration de données en français) est un processus analytique qui consiste à découvrir des modèles ou des relations cachées dans de grandes quantités de données. L'objectif principal du data mining est de transformer des données brutes en informations utiles et exploitables, afin de prendre des décisions éclairées dans différents domaines comme le marketing, la finance, la santé, ou la gestion des risques.

Le Data Mining combine des techniques de statistiques, de mathématiques, de

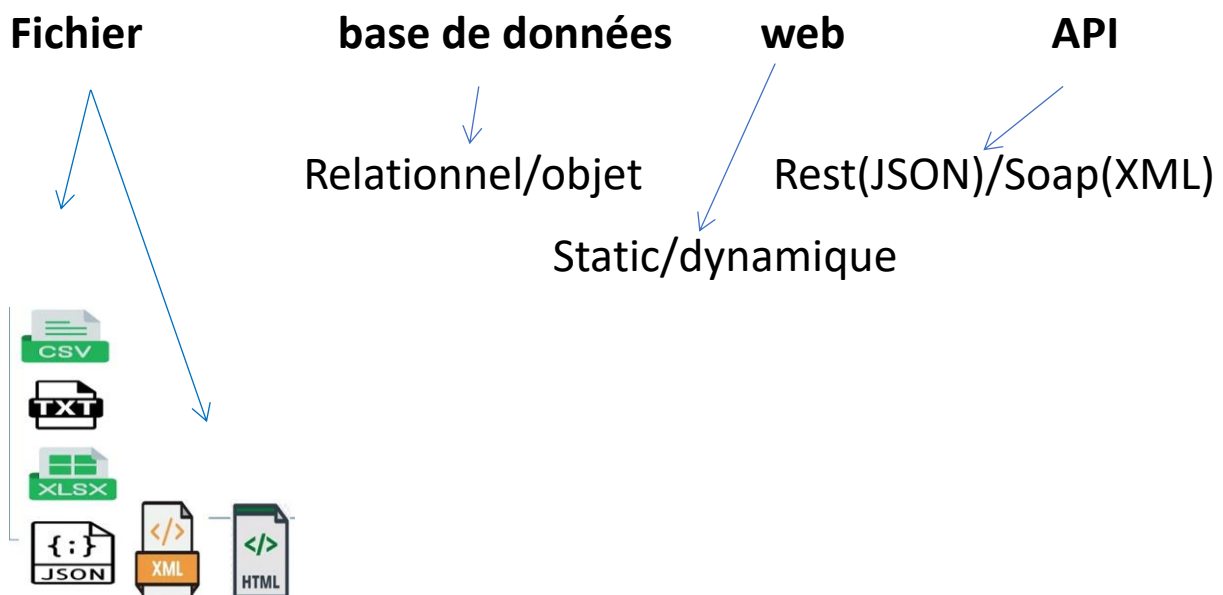
machine Learning, d'intelligence artificielle et de bases de données pour identifier des tendances, des motifs, des corrélations ou des anomalies dans les données.

L'Objectif du DATA MINING :

Les principaux objectifs du data mining sont :

1. Découvrir des modèles cachés : Identifier des motifs récurrents dans les données (par exemple, des habitudes d'achat chez les consommateurs).
2. Prédire les tendances futures : Utiliser des modèles pour prédire des événements futurs (par exemple, prédire les ventes en fonction des tendances passées).
3. Améliorer la prise de décision : Utiliser les insights tirés des données pour prendre des décisions stratégiques ou opérationnelles.

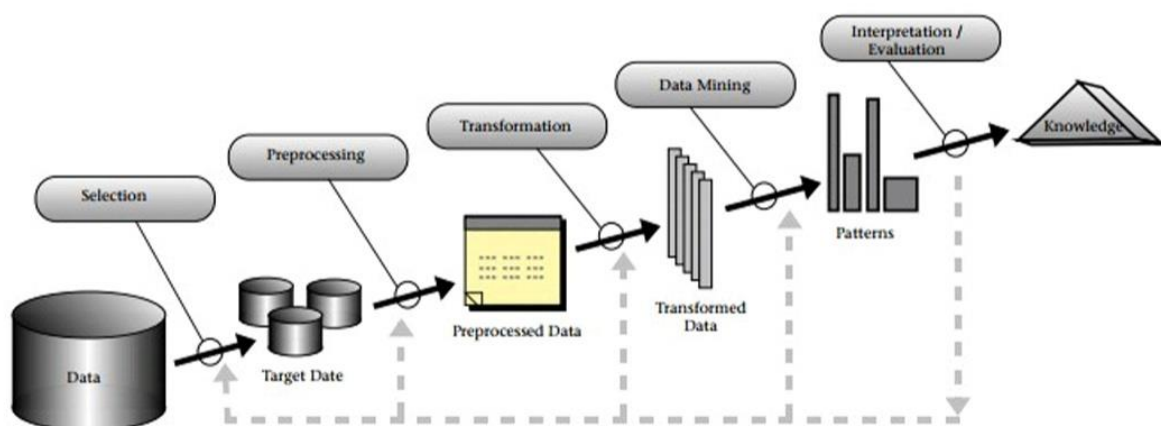
Types des données :



L'importance de python en DATA MINING :

Python est un langage de choix en Data Mining grâce à sa simplicité, sa flexibilité et la richesse de ses bibliothèques. Il permet aux professionnels de l'analyse de données d'effectuer facilement des tâches complexes telles que le nettoyage des données, l'application de modèles de machine Learning, la visualisation des résultats et l'intégration avec d'autres systèmes. Sa popularité et sa large adoption dans l'industrie en font un outil incontournable pour le Data Mining et l'analyse de données.

Pipeline d'exploration de données :



Règle d'association :

La règle d'association est une technique de Data Mining utilisée pour découvrir des relations intéressantes entre des variables dans de grandes bases de données. Elle est principalement utilisée pour identifier des associations ou des patterns fréquents dans des ensembles de données, en particulier dans les domaines du commerce de détail, marketing, ou recommandation de produits.

Une règle d'association est généralement de la forme :

Si A, alors B ($A \Rightarrow B$)

Cela signifie que l'événement A est souvent associé à l'événement B. Par exemple, si un client achète du pain, il achète probablement aussi du beurre

Mesures Associées aux Règles d'Association

Pour évaluer l'efficacité et la pertinence des règles d'association, plusieurs métriques sont utilisées. Les principales mesures sont :

1. Support

Le support mesure la fréquence ou l'importance d'une règle dans l'ensemble des données. Il indique la proportion des transactions qui contiennent les items de la règle, $\text{Support}(A \Rightarrow B) = \text{Nombre de transactions contenant A et B} / \text{Nombre total de transactions}$.

2. Confiance

La confiance mesure la probabilité qu'un item B soit acheté, sachant que l'item A a été acheté. Elle est utilisée pour évaluer la fiabilité de la règle,

$\text{Confiance}(A \Rightarrow B) = \text{Nombre de transactions contenant A et B} / \text{Nombre de transactions contenant A}$.

3. lift

Le lift mesure l'intérêt de la règle en comparant la probabilité de B donnée A avec la probabilité de B indépendamment de A. Un lift supérieur à 1 indique que A et B sont positivement associés (ils se produisent plus souvent ensemble que par hasard).

$\text{Lift}(A \Rightarrow B) = \text{Confiance}(A \Rightarrow B) / \text{Support}(B)$

Application:

Transactions	Articles
1	Pain lait
2	Pain,berre,oeufs,jue
3	Lait, jue, berre,the
4	Pain,lait,jue,berre
5	Pain,lait, jue, the

Reponse:

Element	Transaction	Support
Pain	1,2,4,5	4/5=80%
Lait	1,3,4,5	4/5=80%
Oeuf	2,3,4	3/5=60%
Jue	2,3,4,5	4/5=80%
The	3,5	2/5=40%
Oeuf	1	1/5=20%
Pain,lait	1,4,5	3/5=60%
Pain,lait,berre	4	1/5=20%

Confiance{pain, lait \rightarrow Berre}=

$SUP(Pain,lait,berre)/SUP(Pain, lait)$

$$=0,2/0,6$$

$$=0,33$$

$$=33,33\%$$

$$\text{Lift}\{\text{pain}, \text{lait} \mid \text{berre}\} = 0,33/0,6$$

$$=0,55$$

$$=55\%$$