NOM :zeinebou yahya , NUM :22226

C’est quoi data Mining ?

Data Mining c’est une cherche d'information avancée

C’est quoi Big Data ?

Big data c'est une grand quantite de donne complexes nécessitant des outils spéciaux pour être stockées,traitées et analysées

Relation entre data Mining et Big Data?

Big Data Fournit les donneés et les entils, tandis que data mining utilise des techniques pour en extraire des Connaissances

Relation entre Data Mining et Analyse des donnés

L’analyse des donneés utilise des méthodes statistique pour Comprendre les donnés, tandis que le data mining extrait des modèles cashés à l'aide d’algorithmes avancés

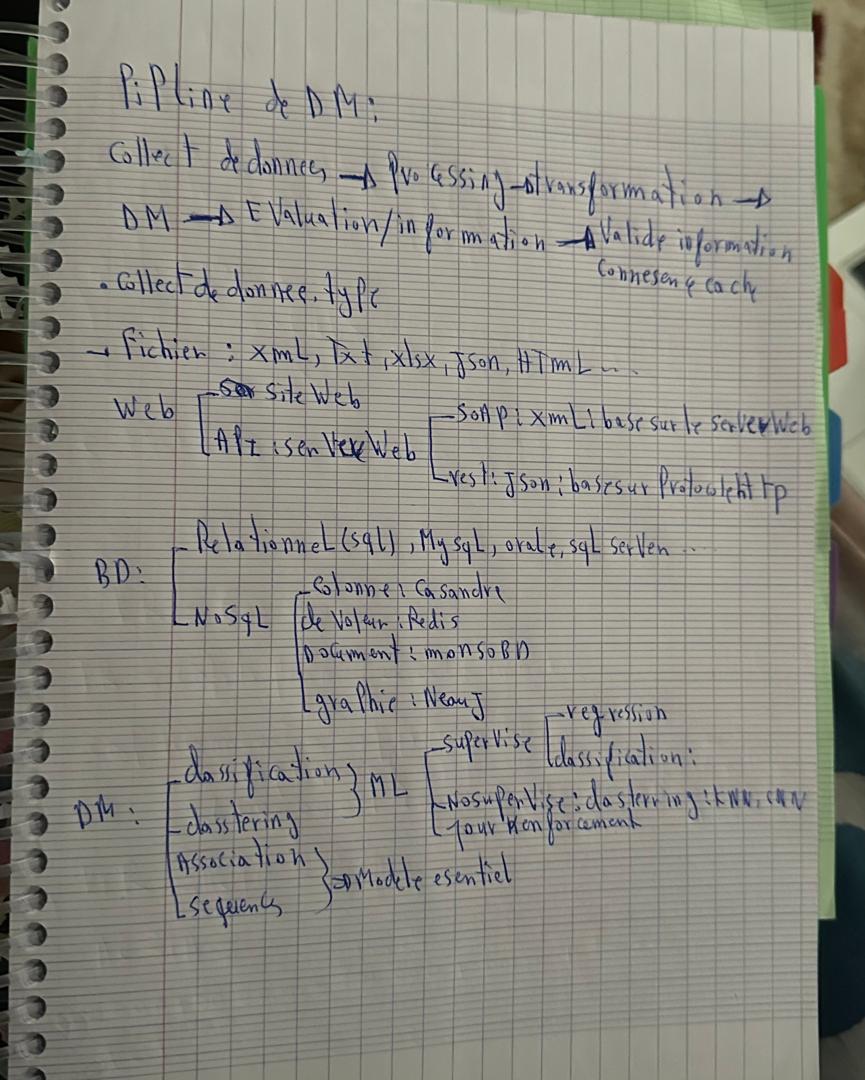
Relation entre Data mining et Intelligences artificielle?

Le DM utilise des techniques d'IA, comme l'apprentissage automatique pour extraire des modéles cachés et améliorer L'analyse des donneés

Méthodes et techniques de DM pour L'exploration des données?

classification, clastering, association regression

Python est un langage facile à apprendre, avec un écosystème riche (Pandas, NumPy, TensorFlow) pour l’analyse de données, l’IA et le Big Data. Il dispose d’une grande communauté, s’intègre bien avec d’autres technologies et est largement utilisé dans l’industrie pour des tâches complexes.



Outils Python pour la collecte de données:

Pandas : Manipulation et analyse des données.

Lecture des fichiers CSV, Excel, SQL.

BeautifulSoup : Analyse HTML et XML pour le scraping web.

Requests : Récupération de données via APIs et services web.

Connecteurs SQL : mysql-connector-python, psycopg2, sqlite3.

Yahoo Finance : Extraction de données financières (prix, historiques).

Pourquoi utiliser différents formats de fichiers ?

Avantages : Compatibilité avec divers outils et applications.

Formats populaires et outils Python associés :

CSV : Format simple et largement supporté (pandas.read\_csv()).

TXT : Format brut pour textes simples (pandas.read\_csv() avec

délimiteurs personnalisés).

XLSX : Données complexes avec feuilles multiples

(pandas.read\_excel()).

JSON : Données structurées, facile à lire (pandas.read\_json()).

XML : Données hiérarchiques (pandas.read\_xml()).

HTML : Extraction des tableaux web (pandas.read\_html()).

Utiliser Pandas pour lire un fichier CSV:

import pandas as pd

df = pd.read\_csv('data.csv’)

print(df.head())

Utiliser Pandas pour lire un fichier Xml:

import pandas as pd

df = pd.read\_xml('data.xml’)

print(df.head())

Utiliser Pandas pour lire un fichier Html:

import pandas as pd

tables= pd.read\_html('https://aws.amazon.com’)

print(tables[0])

Utiliser Pandas pour lire un fichier JSON:

importjson

with open(‘data/data.Json ', ‘r’) as f:

data = json.load(f)

print (data)

Connexion aux bases de données SQL:

import sqlite3

conn = sqlite3.connect('database.db')

query = "SELECT \* FROM table\_name"

df = pd.read\_sql\_query(query, conn)

import sqlite3

REGLES ASSOCIATION :

