

# DATA MINING

## INGSOFT4



BENADJAL.N  
Phd student in computer science  
LAB: MOVEP  
Email: nesrinebenadjel63@gmail.com

2025/2026

# Introduction

Avec l'augmentation exponentielle des volumes de données et l'émergence du paradigme du Big Data, les entreprises sont confrontées à des défis majeurs pour retrouver, extraire et exploiter efficacement l'information enfouie dans ces ensembles massifs. Face à cette complexité, il devient crucial d'adopter des techniques capables de révéler des connaissances pertinentes et de soutenir la prise de décision. Parmi les solutions envisageables, le processus de Data Mining .



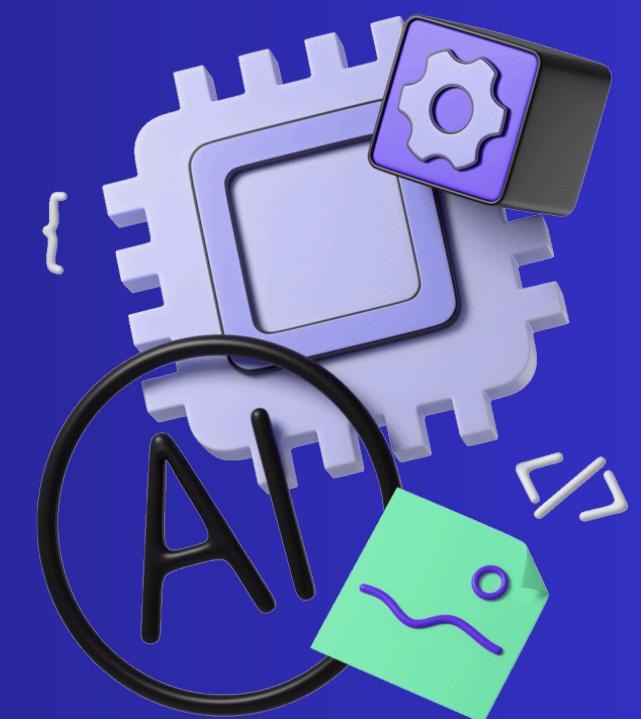
# Qu'est-ce que le Data Mining ?



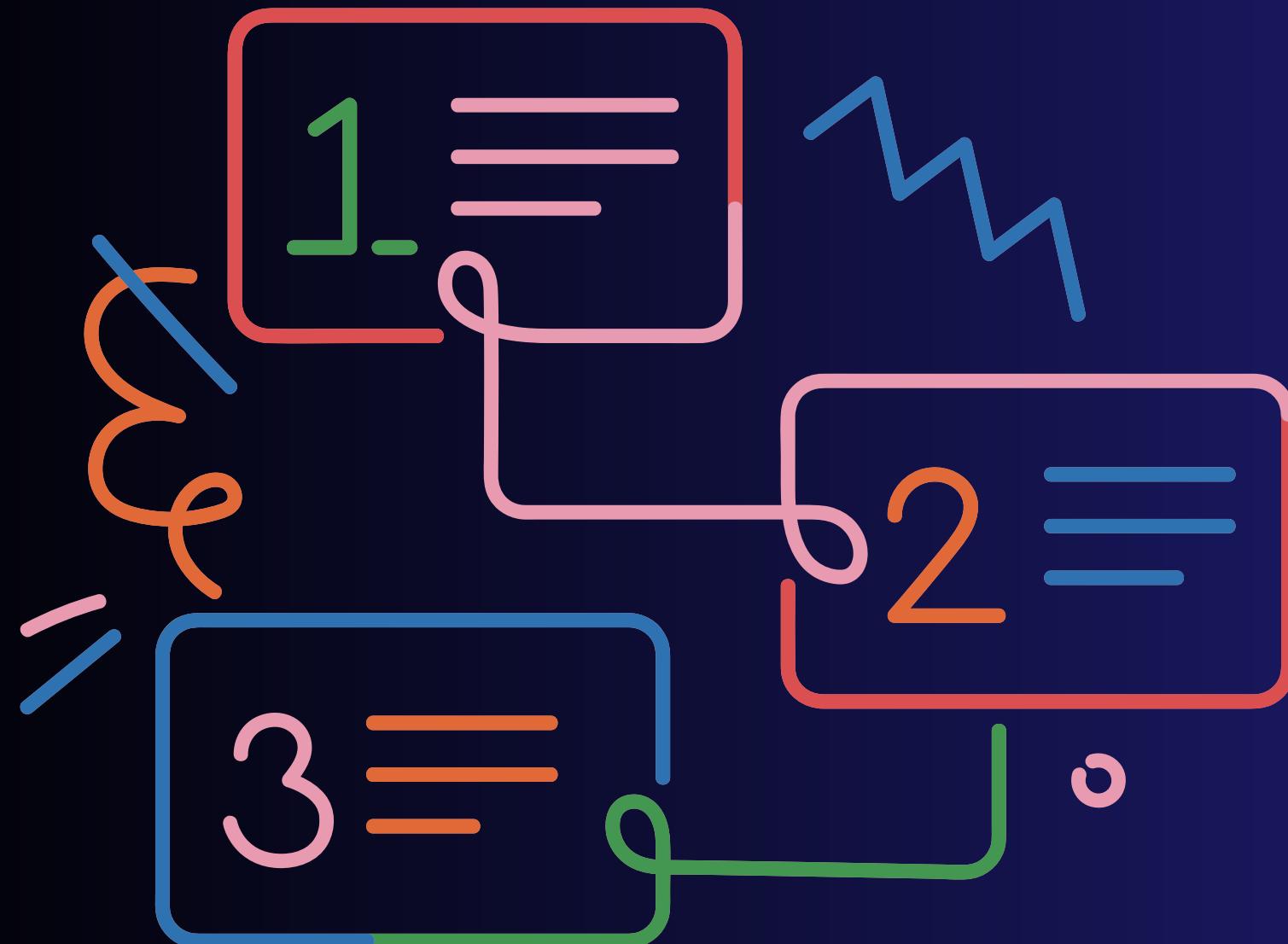
Le Data Mining désigne le processus de recherche et d'extraction de connaissances à partir de grandes quantités de données, dans le but de faciliter la prise de décision.

## Les Méthodes De Data Mining:

1. Association
2. Analyse de sequence
3. Classification
4. Clustering
5. Prédiction



# Les étapes du processus de Data Mining



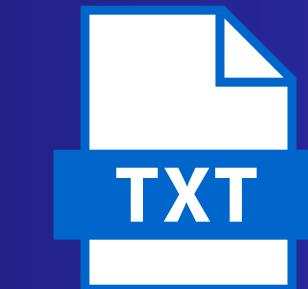
1. Compréhension du Problème.
2. Compréhension des Données.
3. Préparation des Données.
4. Modélisation.
5. Évaluation.
6. Déploiement.

# Les Données

Structuré



Non Structuré



# Ficheie.csv

ATTRIBUTES	Age	Gender	Purchase Frequency	Churn
CUST001	Age	Female	50000	No
CUST002	35	Female	50000	Yes
CUST002	28	60000	8	No
CUST003	42	Female	75000	No
CUST004	19	Male	30000	Yes
CUST005	55	Female	20	No

INSTANCES →

TARGET COLUMN (CLASS)

# DataFrame

Un DataFrame est une structure de données tabulaire en Python via pandas, composé de lignes et colonnes, permet de manipuler, filtrer, analyser des données efficacement.

Les données sont disponibles sous différentes formes et types tels que CSV, table SQL, JSON ou structures Python comme une liste, dictetc. Nous devons convertir tous ces différents formats de données dans un DataFrame afin de pouvoir utiliser les bibliothèques pandas pour analyser ces données efficacement.

	Column Label/ Header				
	0	1	2	3	4
	Name	Age	Marks	Grade	Hobby
0	S1	Joe	20	85.10	A Swimming
1	S2	Nat	21	77.80	B Reading
2	S3	Harry	19	91.54	A Music
3	S4	Sam	20	88.78	A Painting
4	S5	Monica	22	60.55	B Dancing

Diagram illustrating the structure of a DataFrame:

- Index Label:** The row index, labeled 0, 1, 2, 3, 4.
- Column Index:** The column index, labeled 0, 1, 2, 3, 4.
- Row Index:** The row index, labeled 0, 1, 2, 3, 4.
- Column:** A specific column, labeled "Marks".
- Element/ Value/ Entry:** A specific cell value, labeled "60.55".

# Préparation des Données

Dans le cadre de notre premier TP, nous allons nous concentrer sur la préparation des données. Pour cela, nous découvrirons ensemble la bibliothèque Pandas, qui nous permettra de préparer nos données.

