

Analyse de données

Boucherit Ammar
2019-2020

Introduction

Les données ont toujours été importantes

- Les établissements étatiques
Par exemple: **Communes, Wilayattes, Ecoles, Universités ... etc**
- Les entreprises économiques
Par exemple: **Banques, Assurance, Usines, ... etc**
- Les individus (historique médical, arbre généalogique, ... etc)

L'importance était limitée

Pour : **statistique simple, vérification, historique de l'entreprise ... etc**

Archives



Tableaux

Code	Station	Latitude (°N)	Longitude (°O)	Altitude (m)	Période observée	Moyenne annuelle (mm)	CV (%)
2	Agadir	30.26	-9.39	31	1945-2005	228	38.0
3	Casablanca	33.57	-7.38	25	1945-2005	412	32.4
1	Fédakho	33.7	-13.65	6	1945-2005	33.5	68.1
4	El Jachén	33.24	-8.24	10	1945-2005	465	24.5
2	Imzouza	31.91	0.17	1005	1945-2005	127	29.2
6	Ennae	33.32	5.7	2005	1945-2005	579	37.9
7	Khemisat	33.52	5.07	521	1945-2005	374	32.5
11	Marrakech	31.57	-8.08	520	1945-2005	241	27.2
8	Mekrès	33.54	5.31	531	1945-2005	552	27.8
9	Micell	32.58	4.75	1462	1945-2005	205	37.2
10	Dupis	34.59	-9.2	549	1945-2005	414	30.7
11	Béjar	34.99	-5.95	46	1945-2005	536	30.6
12	Safi	32.28	9.24	45	1945-2005	364	37.3
15	Tanger	35.77	-5.81	15	1945-2005	739	31.5
15	Taza	34.25	4.81	560	1945-2005	585	28.7

* CV est le Coefficient de variation en % (CV = (Ecart-type/Moyenne)*100)

Introduction

L'importance a soudainement accrue

Progrès technologique (monde informatique et numérique)

- **L'émergence du phénomène big data**

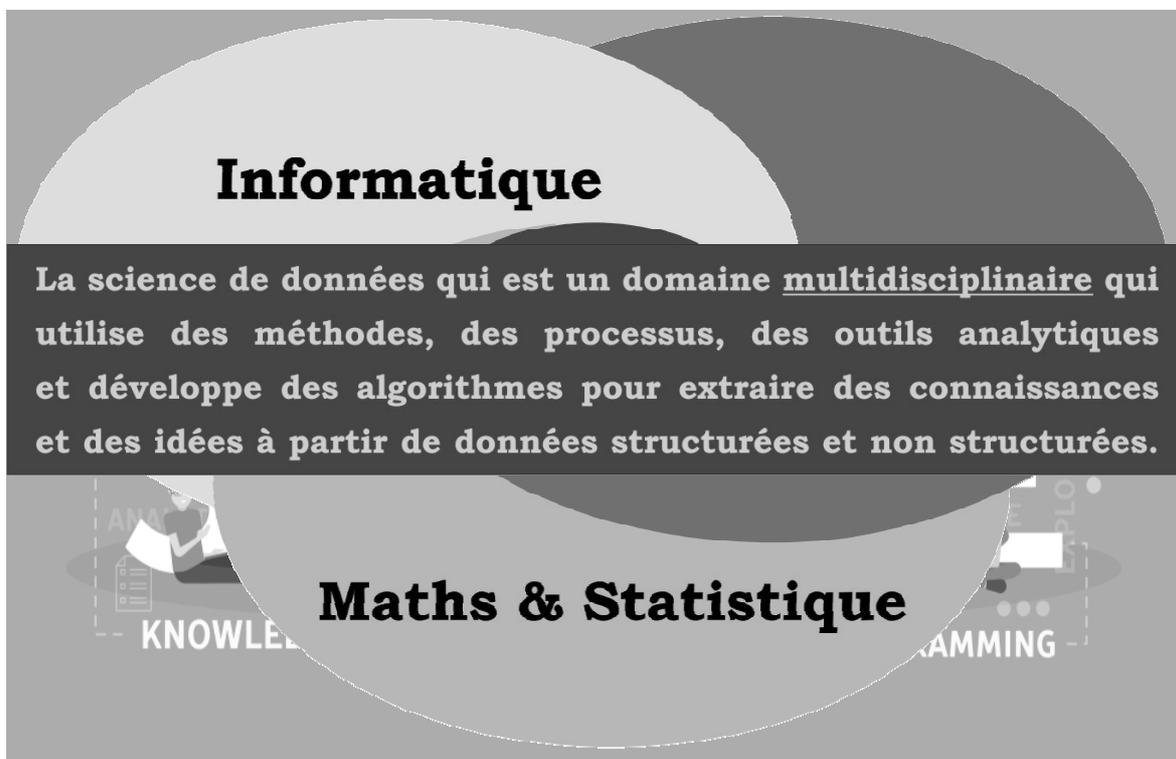
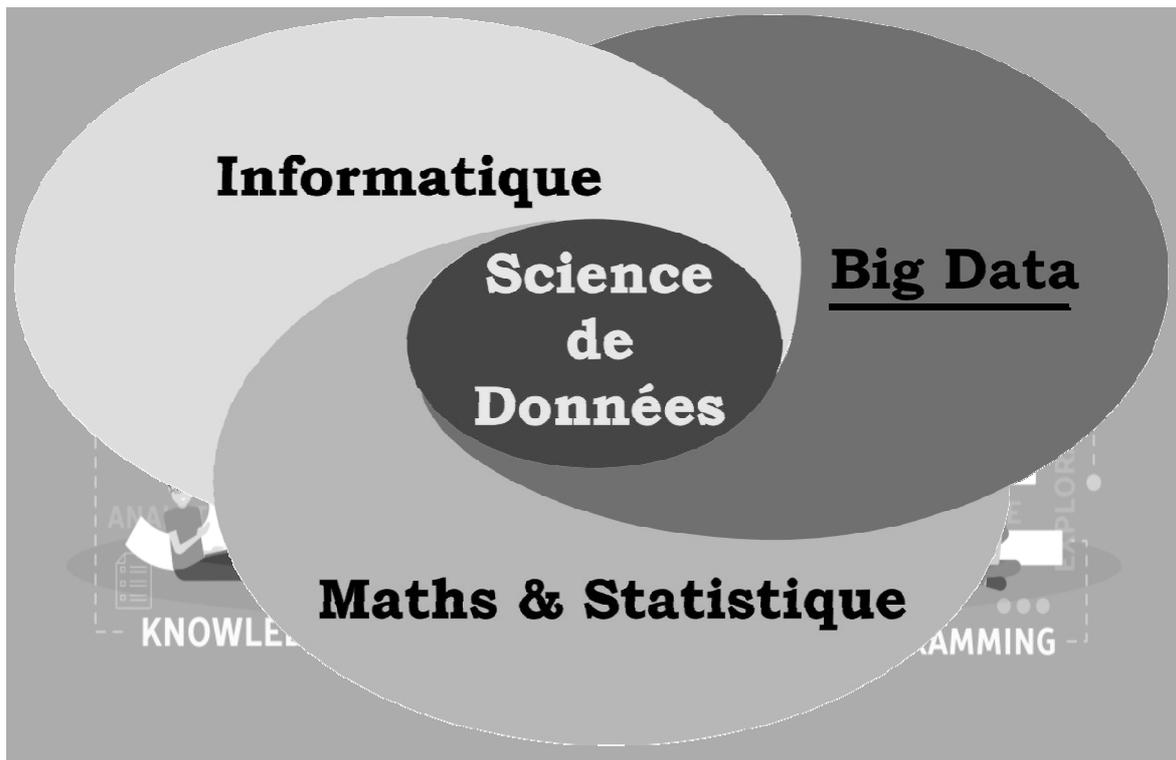
Problème de traitement

- **Les capacités de calcul et de traitement**

↳ Il est devenu nécessaire - dans de nombreux domaines - de tirer parti de cette quantité de données

- processus de décision
- processus de prévision
- meilleure compréhension des phénomènes sous-jacents aux données





Introduction

Domaines d'application

En fait, tous les domaines scientifique qui gèrent de grande quantité de données de type varié (plusieurs dimensions) ont recours à ces approches. On cite à titre d'exemple :

1. marketing
2. météorologie
3. économie
4. écologie
5. assurance
6. banque
7. téléphonie

Contenu du module

- I Introduction
- II Analyse descriptive de données
 - 1 Rappel sur la statistique descriptive
 - 2 Description unidimensionnelle
 - 3 Description bidimensionnelle
- III Analyse Multidimensionnelle Linéaire
 - 1 Analyse Factorielle en Composants Principales
 - 2 Analyse Factorielle des Correspondances (simple)
 - 3 Analyse Factorielle Discriminante
- IV Classification Automatique
 - 1 Eléments de base
 - 2 Classification hiérarchique
 - 3 Classification non hiérarchique