



Extraction des connaissances et BI

Latifa.dekhici@univ-usto.dz

Synonymes

- ▶ L'exploration de données
 - =fouille de données
 - =forage de données
 - = prospection de données
 - =data mining
 - =extraction de connaissances



Définition

- ▶ a pour objet l'extraction d'un savoir ou d'une connaissance (**décision** , découverte.....)
- ▶ à partir de grandes quantités de données (**bases de données hétérogènes**, textes, web...)
- ▶ par des méthodes **automatiques** ou **semi-automatiques**
- ▶ pour un but **opérationnel**

Définition 2

- ▶ l'analyse de bases de données (souvent très grandes) afin de découvrir des relations **insoupçonnées** et de résumer les données d'une manière à la fois compréhensible et utile [Hand et al., 2001]



Insoupçonné

- ▶ Exemples de relations insoupçonnées:
- ▶ Groupe sanguin O → aime Artichauts
- ▶ La Consommation Poivre + Huile est une des causes de cancer.

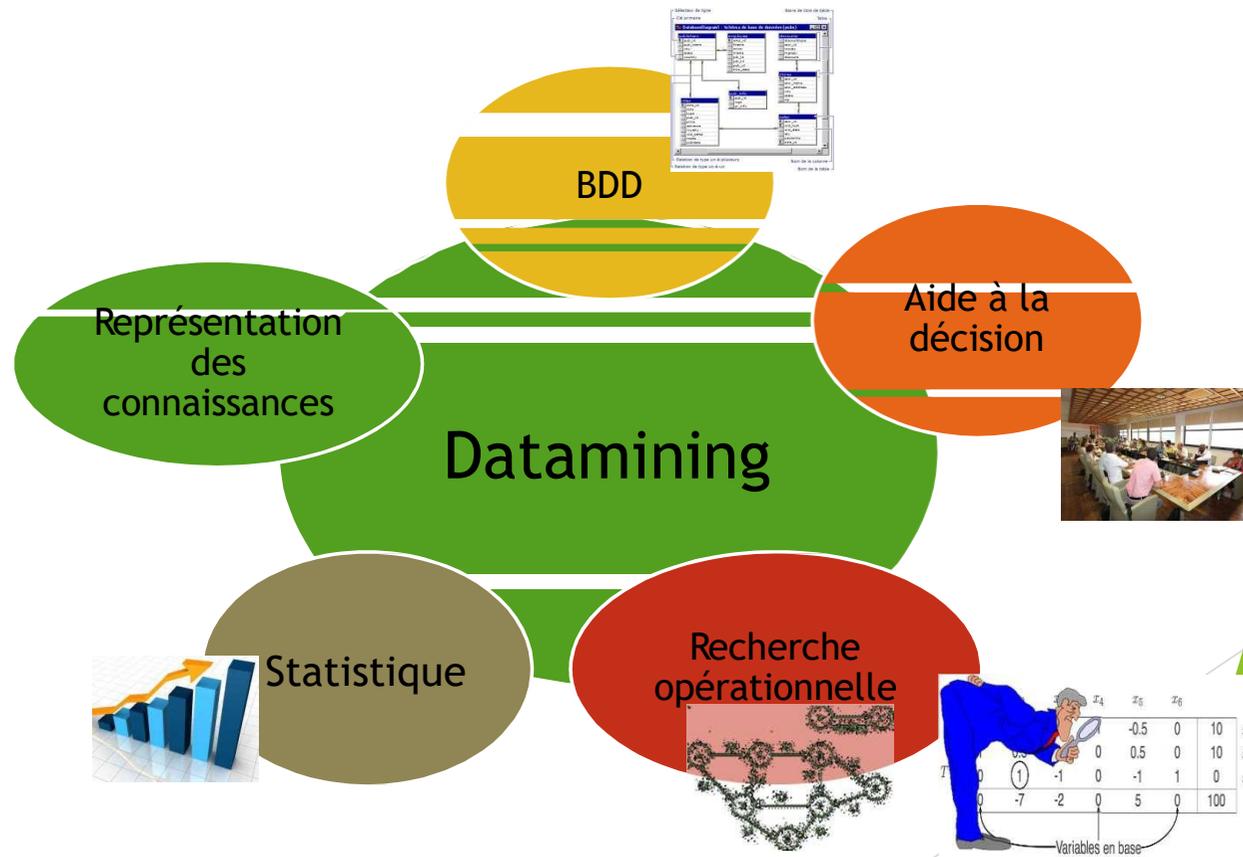


Types

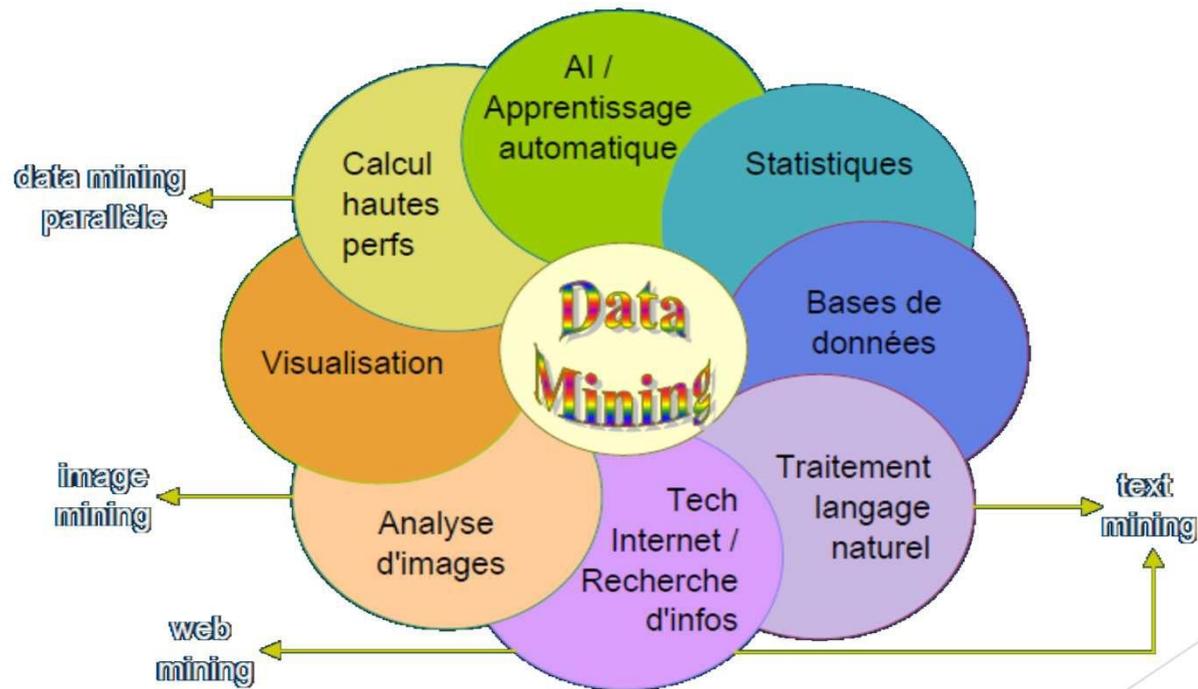
- ▶ Web mining
- ▶ Text mining
- ▶ Image mining
- ▶ Multi media mining



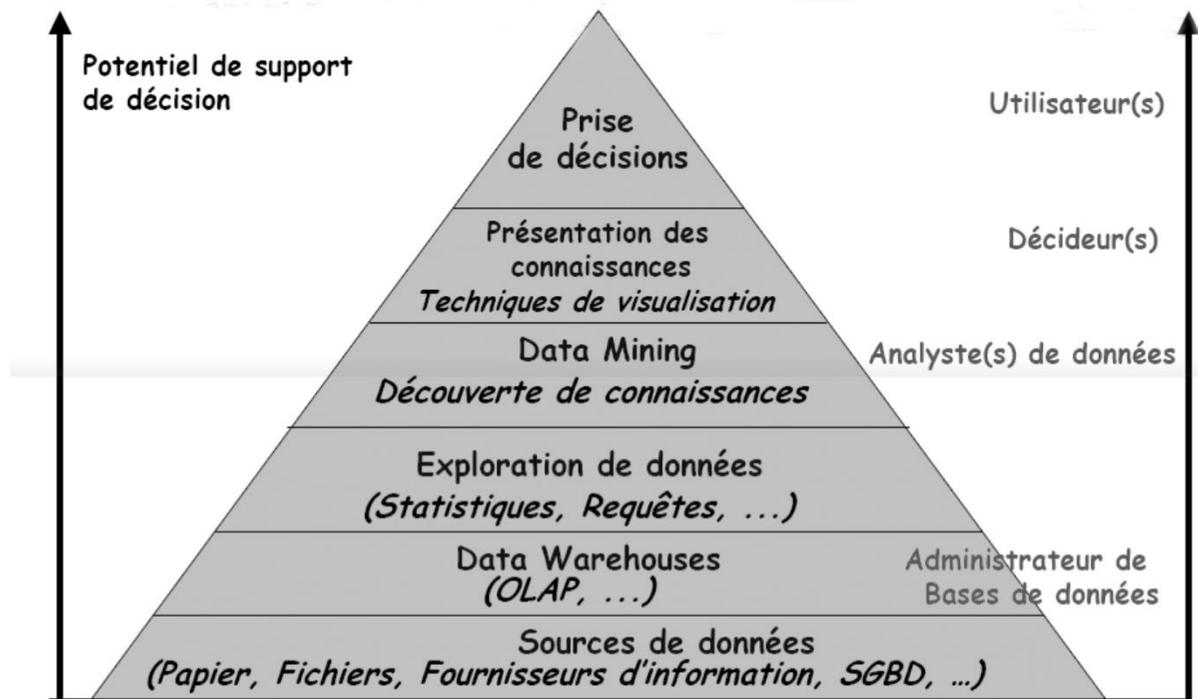
Carrefour



Aujourd'hui, autres branches



Etapes (détails)



Applications

- ▶ Marketing
- ▶ Web
- ▶ Documents
- ▶ Médecine
- ▶ Banque et assurance
- ▶ Police



Marketing ,commerce

► On ne veut plus seulement savoir:

«Combien de clients ont acheté tel produit pendant telle période?»

► Mais:

«Quel est leur profil?»

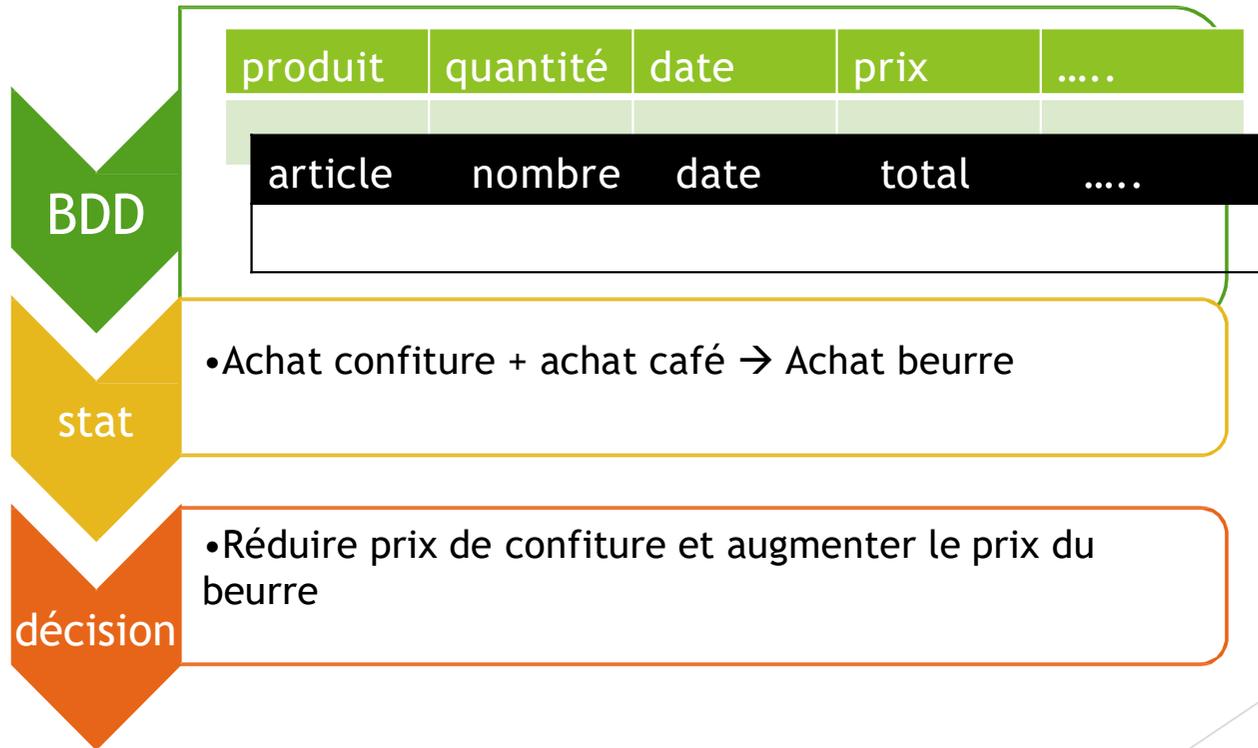
«Quels autres produits les intéresseront?»

«Quand seront-ils intéressés?»

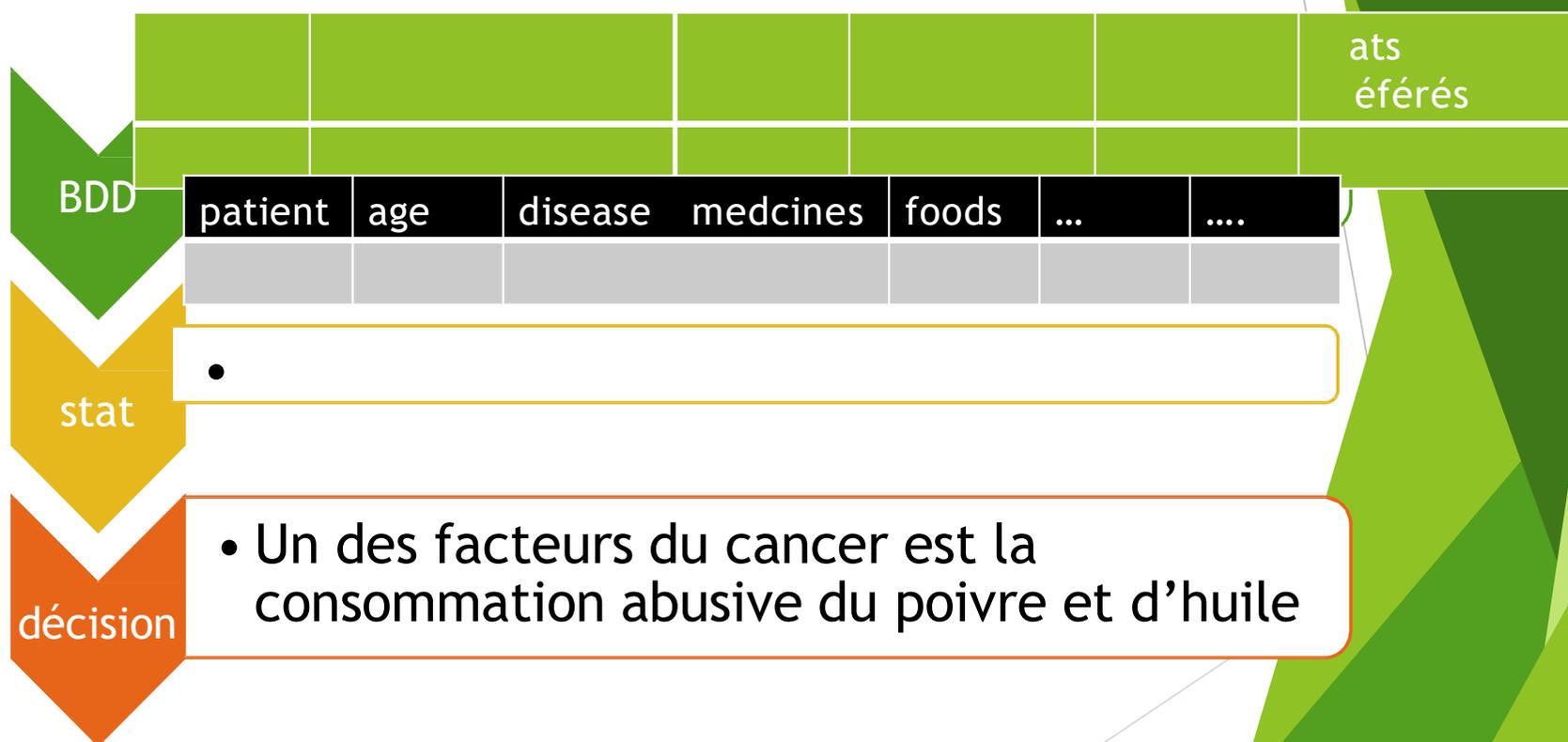
Ce profil permet à lancer des publicités et à industrialiser des produits selon les saisons et les pays



Marketing



Médecine



Médecine

- ▶ Exemple :
- ▶ Le traitement par certains collyres chez la femme enceinte peut entraîner des taches sur les dents de son futur enfant vers l'âge de 6 ans.
- ▶ Donner une cuillère de miel à un nouveau-né peut lui causer le diabète vers 12 ans.

Mailing

► Détection de spam

From : Felix Damians, From .Abidjan Cote D.Ivoire

Hello Dear, Pardon me for not having the pleasure of knowing your mindset before making you this offer and it is utterly confidential and genuine by virtue of its nature. I want someone like you to help me out after i had pray ,then believes that you are a good person and that i can stay with you for the rest of my life , am 24 years old, My late father is a wealthy and successful business man before he died , My mum died when i was a baby, am the only child in my family. And he told me that he used my name to deposit (us dollars 12.5million) in the bank and he seriously advise me to transfer this total money to oversea account for my investment, where i will start my new life and finish my education , Because of this reason, i am soliciting your assistance for the claim and transfer to your bank account for the business. Honestly speaking , i am ready to give you 15percent of this total money for your assistance and with extra 5percent for your expenses on phone call, please reply me now if really serious to help me out so that i can tell you more about my intention and forward to you some of the legal papers after knowing you more better. All about the money are legal and i have all the legal documents and papers of the money money with me which the bank issued to my late father the day of the deposit, Because of the war in our country now the bank manager here has promised to me that they will use their branch corrospodant bank in Europ or Asia.to transfer the money into any account i provided to them that is why i contacted you to help me out in receiveing of the moneymoney into an account over there in your country before i join you over to has rest of mind. Thanks and remain bless with your family as i wait for your urgent reply soonest. Yours Felix Damians.

Web

- ▶ Classification
- ▶ Exemple 1 : le sujet racisme n'a pas de mots clés mais une sémantique
- ▶ Exemple 2 : Le mot magique peut dans une page ne pas référer à la Magie mais à l'informatique, comme Boot Magic
- ▶ Contrôle parental (filtrage)



Web

- ▶ Recherche web
- ▶ Exemple : Une recherche du texte: « informatique Verte » ne doit pas forcément renvoyer des pages contenant cette phrase mais des pages ayant une relation avec « l'optimisation du hardware par exemple etc.»

Police data mining

- ▶ Détecter la signature d'un criminel

Toutes ses victimes portaient des lunettes et avaient des voitures grises

- ▶ Découvrir l'identité réelle d'une personne par son compte Facebook

Exp: data detective



Techniques de data mining

Parmi les techniques les plus simples:

- l'algorithme à priori basé sur les règles
- et les arbres de décision.

Pour renforcer ces algorithmes des techniques sont ajoutées tel que les réseaux de neurones

Règle d'association

A,B, C → D

Exemple:

Achat PC → achat sac à dos



Mesures des règles d'association

Les **Mesures objectives** qui souvent favorisent les items fréquents et on peut citer le *support* et *confiance*.

support, s , Le support de l'ensemble XYZ correspond à l'occurrence des trois éléments en même temps ou la probabilité qu'une transaction contienne {X, Y, Z}.

confiance, c , probabilité conditionnelle qu'une transaction qui contient {X, Y} contienne aussi Z.

Confiance($x, y \rightarrow z$) = $\text{support}(X, Y, Z) / \text{support}(X, Y)$

Exemple de support et de confiance

L'apparition de X =toux sèche, Y=respiration difficile et Z=Maladie COVID-19 apparaissent dans 20 cas d'une base de 70 bases.

support(X,Y,Z)=20/70=28%.

L'apparition de X =toux sèche Y=respiration difficile dans les bases et pour 40 cas parmi la base des patients.

Alors support(X,Y)=40/70=56%.

La confiance de la règle d'association(X =toux sèche et Y=respiration difficile → Z=Maladie COVID-19) confiance (x, y → z)=20/40=50%.

Seuils

Certaines décisions peuvent être éliminées en exigeant **un seuil pour les supports et la confiance minimale**. Plus la confiance est élevée, meilleure est la règle d'association.