

Knowledge Building

*Transformer un large volume d'informations en connaissances directement exploitables pour prendre des décisions essentielles, voire vitales, tel est l'enjeu du **Knowledge building**.*

L'objectif **du Knowledge building** –incluant notamment les techniques de modélisation et de data mining (ou fouille de données) - est de fournir les clés permettant la transformation d'un **ensemble de renseignements** en **connaissances** afin d'en retirer un **avantage déterminant**.

Cette approche va au-delà de l'analyse statistique de données, ou même des systèmes dits « décisionnels », dont l'objectif s'arrête essentiellement à l'extraction et à la présentation d'informations (faits).

En effet le data mining, en explorant une base de données, va s'attacher à y détecter un ensemble de règles, d'associations, de motifs (patterns) ou de tendances, qui permettront ensuite de prévoir des faits avec un haut niveau de probabilité.

Il pourra donc servir, en particulier, à atteindre deux objectifs :

1. A partir de données caractérisant un contexte ou une situation, le data mining permettra la modélisation d'une conséquence (un fait à venir). Quelques exemples :
 - **Marketing** Connaissant les caractéristiques de tel ou tel prospect, quel est le niveau de probabilité que celui-ci soit intéressé par mon offre commerciale ?.
 - **Industriel** : Connaissant la composition chimique et les caractéristiques physiques et mécaniques de tel ou tel bitume (viscosité, teneur en soufre, etc.), quelles seront les caractéristiques mécaniques d'un enrobé fabriqué à partir de celui-ci ?
 - **Sciences de la vie**: Connaissant les caractéristiques d'une plante ou d'arbuste, quelle sera sa faculté d'acclimatation à telle ou telle région? Quel sera son impact, à terme, sur l'environnement ?
2. Déterminer les caractéristiques du système optimal qui maximisera telle ou telle fonction (détermination de l'optimum). Exemples :
 - **Marketing** : Parmi la population, quelle est la clientèle que je dois cibler en priorité pour optimiser ma campagne marketing ?
 - **Industriel** : Quel est l'assemblage optimal de quels bitumes pour atteindre l'objectif que je me suis fixé en termes de qualité mécanique d'enrobé bitumineux ?
 - **Sciences de la vie**: Quelle serait la culture la mieux adaptée à telle ou telle région ?

Par conséquent, le data mining permet de valoriser des données afin d'en dégager un avantage concurrentiel majeur

Qui se sert du data mining ?

Aujourd'hui, le data mining est souvent utilisé dans les secteurs de la Finance, du Marketing et du Commerce, mais il peut potentiellement concerner un grand nombre d'autres secteurs.

Voici les résultats d'une étude des secteurs exploitant le data mining, publiée par la communauté KDNuggets.com (2010) :

Branches et domaines dans lesquels est utilisée l'exploration des données (%).

GRC/ analyse de la consommation	26,8	Banque	19,2	Santé, ressources humaines	13,1
Détection de fraude	12,7	Finance	11,3	Marketing direct, collecte de fonds	11,3
Télécommunication	10,8	Assurance	10,3	Sciences	10,3
Éducation	9,9	Publicité	9,9	Web	8,9
Médical	8,0	Produits manufacturés	8,0	Commerce de détail	8,0

Quelles données utiliser pour un projet de data mining?

La phase la plus critique d'un projet d'exploration de données est celle de la collecte et de la préparation des données qui seront ultérieurement exploitées. Cette phase doit impérativement commencer par poser la bonne question et formaliser le problème à résoudre, ce qui suppose une compréhension claire et précise de la problématique « métier ».

Le volume de données sera en général très important, car le système devra passer par une phase d'apprentissage aussi complet que possible. Se contenter de faibles quantités de données pour formaliser un problème complexe conduit en général à des conclusions erronées.

Là où les études statistiques et les méthodes classiques d'analyse de données portaient sur des populations de dizaines de milliers d'individus caractérisés par des dizaines de variables, les algorithmes de data mining pourront compiler des bases de millions d'individus, comportant plusieurs centaines de variables, numériques ou non.

Les méthodes du data mining

De nombreuses méthodes et algorithmes sont disponibles pour réaliser un projet de data mining.

On peut citer deux grandes familles d'algorithmes : les méthodes dites **descriptives**, qui aident à comprendre les informations et à dégager ainsi des groupes homogènes et des normes de comportement et les méthodes dites **prédictives** qui, comme leur nom le suggère, s'attachent à expliquer ou à prévoir des phénomènes observables et effectivement mesurés.

Résoudre un problème par un processus d'exploration de données impose généralement l'utilisation d'un grand nombre de méthodes et d'algorithmes différents.

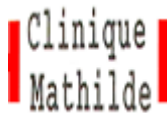
Citons parmi celle-ci la méthode des **Arbres de décision**, assez courante car lisible et rapide à exécuter, ou encore celle des **Réseaux neuronaux**, inspirée du fonctionnement des neurones biologiques, qui nécessite en général une longue phase d'apprentissage mais qui permet de réaliser des classifications rapides, notamment pour la reconnaissance de motifs (Reconnaissance optique de caractères par exemple).

Que peut apporter Sonoma ?

Sonoma dispose d'une compétence avérée dans le domaine de l'exploration de données, acquise au travers de projets complexes menés dans des contextes variés : Parmi ceux-ci, citons en particulier:



Conception et réalisation d'un projet d'exploration de données destiné à optimiser la qualité et le coût de fabrication de bitumes, par analyse des caractéristiques d'échantillons de pétrole brut provenant de différents gisements.



Exploration de données radiologiques dans le but d'améliorer la prévention et le diagnostic de certaines pathologies vasculaires.



Exploration en temps réel d'un ensemble d'informations issues d'équipements d'acquisition (capteurs, radars, etc.) pour améliorer la prévention et le traitement des risques affectant la sécurité d'un site sensible.

Notre expérience nous permet d'aborder des contextes variés et de conseiller efficacement nos clients sur la ou les méthode(s) à utiliser pour résoudre un problème ou répondre à une question, à l'aide de l'exploration de données

Nous prenons ensuite en charge la réalisation du projet jusqu'à la production des résultats et des conclusions, dans un format adapté (rapports, tableaux, graphiques 2D, 3D, etc.)

Pour nous contacter :

SONOMA

35, rue du pré de la bataille

76000 Rouen

Tel : 02 35 60 00 42

E-mail : contact@sonoma.fr