

## Le Data Mining au service du CRM

### Démarche intellectuelle du data mining

Sur le plan technique la démarche s'expose plus facilement en anglais :

- « pattern recognition » : on identifie les motifs récurrents, formes répétitives, signes d'organisation, structures cachés ou modèles enfouis dans ses données
- « data driven hypothesis generation » : on les utilise pour faire émerger de nouvelles hypothèses
- « value analysis » : on voit ce qu'elles peuvent nous apporter

### Quelle est la différence avec l'approche statistique ?

La démarche statistique relève généralement d'une démarche scientifique où l'on s'emploie à **valider des hypothèses par les expériences** que l'on conçoit.

A l'opposé de cela, considérée seul, le data mining suit une démarche NON scientifique pour trois raisons :

- on part le plus souvent sans hypothèse.
- souvent les données dont émergent les hypothèses existent déjà au moment l'on intervient : on ne contrôle ni leur choix, ni leur recueil, ni leur qualité
- ce que l'on observe est difficilement reproductible dans la mesure où les phénomènes observés évoluent rapidement (nouvelles offres, nouveaux concurrents, évolutions des comportements, etc.)

Un projet d'analyse de données représente en général 3 étapes :

- obtention de données significatives
- analyse
- industrialisation de processus sur la base des résultats de l'analyse

La première étape consiste en la réalisation d'un système de collecte ou de fabrication de données lorsque celles-ci n'existent pas.

La mise en place d'un **système d'un questionnaire en ligne de satisfaction** client est, par exemple une méthode très efficace pour obtenir des informations précise par profil client

La deuxième étape est celle qui met en jeu le cœur du datamining dans l'optique par exemple de déterminer des relations, de réaliser des prédictions, de faire émerger une ou plusieurs classifications...

Il reste enfin à exploiter ces découvertes par le développement de traitements informatiques qui automatisent la collecte de données en flux continu, qui appliquent les modèles déterminés à l'étape précédente et restituent les résultats opérationnels à un client informatique ou humain.

Les traitements informatiques sont parfois chargés d'introduire une boucle réactive pour automatiser la mise à jour du modèle dans son évolution lente (évolution des catégories d'une classification, de paramètres d'un système prédictif...).

En réalité, ce n'est pas aussi simple, l'extraction de connaissances à partir de bases de données possède des particularités qui sont loin d'être négligeables :

(1) des techniques d'analyse qui ne sont pas dans la culture des statisticiens, en provenance de l'apprentissage automatique (Intelligence artificielle) et des bases de données ;

(2) l'extraction de connaissances est intégrée dans le schéma organisationnel de l'entreprise. Ainsi, les données ne sont plus issues d'enquêtes ou de sondages mais proviennent d'entrepôts construits sciemment pour une exploitation aux fins d'analyse, le DATAWAREHOUSE. D'une part, une réorganisation du flux de données au sein de l'entreprise devient nécessaire (l'enchaînement des bases de production, le Datawarehouse et les Datamarts); d'autre part, la capacité des méthodes statistiques à traiter de gros volumes devient un élément clé ;

(3) enfin, dernier élément important, le traitement des données sort de plus en plus des sentiers battus en traitant, non plus seulement des fichiers plats "individus x variables", mais également des données sous forme non structurée, le texte, depuis un bon moment déjà, mais aussi les images et la vidéo. Cette orientation attribue une place primordiale à l'appréhension et la préparation des données.

#### Application de ces nouvelles technologies dans l'entreprise.

Comme le suggère la définition, le datamining est une spécialité transverse : elle regroupe un ensemble de théories et d'algorithmes ouverts à tout domaine métier susceptible de drainer une masse de données.

**Qualitelis** s'intéresse particulièrement au développement du Data Mining dans l'étude comportemental des clients et dans le domaine du CRM.

- Profiling client pour un service marketing d'une entreprise de distribution
- Profiling produit pour améliorer le cross-selling dans la grande distribution
- Profiling client pour anticiper les besoins des clients et dépasser leurs attentes.

Généralement, l'utilisateur qui veut bénéficier d'une analyse prédictive va :

- extraire ses données de son environnement habituel,
- les importer dans l'outil de data mining,
- demander à un informaticien connaissant l'outil et son langage de programmation de bien vouloir procéder à l'analyse,
- lui demander aussi de bien vouloir exporter les résultats vers son environnement habituel, à chaque fois que ceux-ci évoluent.

Ce n'est plus ce que les utilisateurs veulent faire. Quand ils écrivent un document dans Word, ils ne veulent ni ouvrir un dictionnaire, ni lancer une application externe : ils veulent que Word leur suggèrent la bonne orthographe automatiquement.

De la même façon, un commercial plongé dans son application de CRM (Customer Relation Management) voudrait non seulement connaître la combinaison de caractéristiques qui identifie le mieux ses prospects à fort potentiels, mais directement leur envoyer un mailing ciblé, par profil de client sans quitter son application.

Dans cette optique, les composants de data mining présents dans les services offerts par « Qualitelis » seront conçus :

- pour fournir une analyse prédictive à la fois
  - o stratégique : comment évolue la clientèle ciblée ? quelle valeur perçoit-elle dans l'offre de l'entreprise ?
  - o et tactique : quels est la liste des prospects à cibler ? par quels moyens ? à quel moment ? avec quel message ?
- pour les entreprises qui ne disposent pas forcément d'un département statistique consacré aux analyses exploratoires et prédictives,
- pour être programmés via des langages et des outils généralistes,
- pour accroître l'intérêt des outils OLAP et reporting,
- pour se fondre de façon transparente dans les applications déjà déployées dans l'entreprise.

## 2 exemples d'applications :

a) **Un opérateur télécom** souhaite retenir les abonnés tentés par la concurrence

Il pourra limiter l'attrition de sa base d'abonnés :

- en découpant sa base de clients en groupes utiles tels que « *les clients qui m'intéressent* » (ie : susceptibles d'être rentables) et *les autres* par l'analyse des clusters correspondants,
- en identifiant les caractéristiques propres aux clients à risque (tranche d'âge, tranche de revenus, évolution récente de la consommation téléphonique, etc.) et la façon dont celles-ci se contribuent à expliquer les résiliations constatées sur une période plus ou moins récente,
- en faisant bénéficier les clients à risques de promotions spéciales susceptibles de les faire changer d'avis.

b) **Une grande surface** souhaite identifier les meilleurs prospects pour son nouveau rayon cuisine équipée

Elle dispose de deux sources d'information :

- des données recueillies via la carte de fidélité : âge, revenus, lieu d'habitation, etc.
- des données d'usage recueillies lors du passage en caisse : heures de visites, achats réalisés, achats réalisés ensemble lors d'une même visite, achats réalisés lors de visites successives, évolution du ticket moyen, etc.

Leur analyse automatique permettra d'extraire :

- les meilleures caractéristiques prédictives, telles que « femmes entre 30 et 50 ans habitant le département »,
- les associations d'achat les plus révélatrices, telles que « l'achat d'une cuisine est souvent précédé par celui d'un rideau de douche » (dans ce cas l'acheteur vient probablement d'intégrer un appartement vide, qu'il aura peut-être envie d'équiper un peu plus tard).

Pour exploiter l'information découverte, la grande surface pourra entre autre :

- réaliser des publipostages ciblés
- installer le rayon cuisine près du rayon salle de bains et suggérer aux vendeurs d'aborder toute pimpante trentenaire rôdant près des rideaux de douche !

**Quelques points de réflexions concernant les applications du Data Mining en grande distribution :**

Les premiers essais de fouilles de données se firent historiquement sur des dépouillements de millions de tickets de caisse de supermarché tels que mémorisés par les caisses enregistreuses.

À l'origine de la vulgarisation des méthodes et des algorithmes de la fouille de données (en l'occurrence, l'analyse des associations), il y aurait eu la mise en évidence par certaines grandes surfaces aux US d'une corrélation très forte entre l'achat de couches pour bébés et de bière le samedi après-midi<sup>2</sup>.

Les analystes s'aperçurent alors qu'il s'agissait des *messieurs* envoyés au magasin par leur *compagne* pour acheter les volumineux paquets de couches pour *bébé*. Les rayons furent donc réorganisés pour présenter côte à côte les couches et les packs de bière ... dont les ventes grimperent en flèche ! Cette image pourrait être un exemple du *retour sur investissement* des travaux de data mining et plus généralement d'informatique décisionnelle.

D'autres analyses, en particulier d'indices précurseurs de modification des habitudes d'achat, se montrèrent d'un plus grand intérêt dans la pratique.

Un autre exemple souvent cité est le suivant : si on baisse le prix du Coca-Cola de 5%, on va par exemple en augmenter les ventes de 15%, ce que l'on savait sans data mining. Mais le data mining révèle l'élément inattendu (bien qu'évident *a posteriori*), à savoir que les ventes des *cacahuètes* vont augmenter dans une proportion voisine (sans doute suite à l'association d'idées : « Puisque j'achète du Coca, il me faut aussi des cacahuètes » ; le data mining ne fait pas d'hypothèse sur le sujet). Si la marge sur le Coca-Cola est relativement faible, et celle sur les cacahuètes importante, la conclusion s'impose d'elle-même : *baisser le prix du Coca-Cola est un moyen de vendre davantage de cacahuètes*.

Par le côté quasi-déterministe qu'elle présente du comportement du consommateur, cette approche provoque souvent une réaction de scepticisme, mais dès lors que les résultats de sa mise en œuvre sont au rendez-vous, ce qui est quelquefois le cas, son adoption ne pose ensuite pas de problème particulier.