

## Cinq Mythes sur le Data Mining



Mars 2006

**Michel Bruley**  
Directeur Marketing Teradata France

Pour plus d'information sur ce livre blanc veuillez contacter :  
[michel.bruley@ncr.com](mailto:michel.bruley@ncr.com)

Un livre blanc de Teradata

**Teradata**  
a division of  NCR

## Cinq mythes sur le data mining

---

De nombreuses entreprises florissantes ont découvert que les bruits qui circulent à propos du data mining ne se résumaient finalement qu'à de simples mythes. Plutôt que de céder à ces sirènes, ceux qui ont su voir plus loin y ont gagné un formidable avantage concurrentiel en utilisant le data mining pour résoudre des problèmes d'entreprise complexes et voir augmenter leur rentabilité.

Au fur et à mesure que les compagnies de télécommunications poursuivent ce petit « plus » qui permettra de relancer le revenu moyen par utilisateur (ARPU) et d'améliorer la rentabilité et l'efficacité de leurs opérations, le data mining et la transformation de données en informations directement utilisables tendent à se répandre toujours plus largement. Il reste néanmoins beaucoup à faire. Sous la pression accrue de la concurrence et des contraintes réglementaires, les compagnies de télécommunications cherchent à offrir un meilleur service pour un coût réduit. Fournir un meilleur service ne se résume pas à offrir une qualité de réseau optimale : cela implique aussi d'être en mesure de comprendre réellement les besoins des utilisateurs et d'y répondre à l'aide de services rentables. De même, réduire les coûts ne signifie pas simplement diminuer les prix. Dans la conjoncture économique actuelle, la pression des tarifs agit souvent comme un catalyseur déclenchant l'examen des systèmes, des process et des coûts internes, ainsi que la recherche de l'excellence opérationnelle. Offrir un service meilleur et plus rentable, d'une part, et atteindre l'excellence opérationnelle, d'autre part, sont pour les entreprises deux objectifs pour lesquels la technologie du data mining peut constituer un atout précieux.

C'est d'ailleurs une technologie de data mining de pointe qui a convaincu les magasins Wal-Mart installés à la campagne de miser sur une version spécifique de Spam pour la saison de la chasse. Allez-y. Riez. Il n'empêche que la vente de Spamouflage-Spam dans des boîtes en tenue de camouflage s'est révélée un énorme succès. Au-delà d'une idée simplement séduisante, Spamouflage a aidé Wal-Mart à tirer des revenus supplémentaires de ses clients existants et a démontré à quel point ce distributeur comprenait les gens à qui il s'adresse.

Bien qu'étranger au monde des télécommunications, cet exemple n'en reste pas moins significatif de la façon dont des entreprises leaders sur leur marché peuvent transformer des données en informations, puis en actions et, au bout du compte, en bénéfices.

### Mais qu'est-ce donc que le data mining ?

Le data mining est un puissant outil d'analyse qui permet à des responsables d'entreprises d'aller plus loin que la simple description des comportements passés d'une clientèle et de prédire l'avenir. Il permet de déceler dans les comportements mystérieux des clients des règles qui les sous-tendent. Grâce à ces découvertes, il est possible d'augmenter les revenus, de réduire les dépenses, d'identifier des opportunités commerciales et donc de bénéficier d'avantages concurrentiels certains.

Le développement de mythes autour du data mining s'explique en partie par le fait que le concept reste encore flou pour bien des gens. À la base, le data mining se définit comme un jeu de techniques mathématiques complexes servant à découvrir et à interpréter des schémas jusqu'ici inconnus dans un ensemble de données détaillées. Depuis le milieu des années 1980, période à partir de laquelle le data mining a commencé à sortir du champ de la recherche universitaire, médicale et scientifique, ces techniques ont été appliquées avec une grande efficacité dans le domaine des télécommunications, de la distribution, des banques, des assurances, des transports et de l'hôtellerie.

La réputation d'outil analytique du data mining l'a souvent fait confondre avec la technique de traitement analytique en ligne nommée OLAP. L'OLAP est une technique analytique précieuse lorsqu'on l'applique à l'analyse d'opérations commerciales afin d'en tirer une perspective historique des événements. Imaginons par exemple qu'un directeur du marketing veuille comprendre pourquoi les ventes ont chuté dans une région donnée. Les outils d'OLAP lui permettent d'interroger la base selon de multiples critères, qu'il s'agisse des ventes par segment, par produit, par service ou de leur évolution. En examinant l'historique des données sous plusieurs angles, ce responsable sera en mesure d'identifier les

causes (point de vente, produits ou période) qui ont pesé sur les ventes.

Le data mining s'intéresse à un ordre de problèmes différent. Il peut servir à prédire des événements futurs, comme les ventes du mois suivant en fonction des promotions ou le type de consommateur qui sera le plus sensible à une vente promotionnelle. La façon dont un certain nombre d'entreprises utilisent déjà permet de dissiper les cinq grands mythes qui planent sur le data mining.

### Mythe numéro un : le data mining pourrait instantanément prévenir l'avenir, à la manière d'une boule de cristal.

Le data mining n'est ni une boule de cristal ni une technologie capable de faire apparaître des réponses comme par magie en appuyant sur un bouton. C'est un processus qui comporte plusieurs étapes : définition du problème de l'entreprise, exploration et préparation des données, développement d'un modèle et déploiement des connaissances obtenues. En général, les entreprises consacrent l'essentiel du temps nécessaire à pré-traiter et à préparer les données afin d'obtenir des Données nettoyées, cohérentes et convenablement combinées pour fournir les renseignements fiables qu'elles en attendent. La qualité des données est au cœur du data mining et un data mining réussi exige des données reflétant fidèlement l'activité de l'entreprise. Dans le cas des télécommunications, les opérateurs doivent comprendre que la puissance du data mining réside dans sa capacité à relever des défis spécifiques de nature prédictive ou descriptive. Il peut s'agir de :

- Segmenter la clientèle,
- Prédire la propension des clients à acheter (ou à changer de prestataire),
- Détecter la fraude,
- Augmenter l'efficacité de l'organisation (par exemple en optimisant l'assurance de revenus et les performances réseau).

Les entreprises de télécommunications qui comprennent comment fonctionne le processus en retirent des résultats indéniables. Tel opérateur européen de téléphonie mobile connaît par exemple la valeur de ses clients. Pas seulement les revenus qu'il peut en tirer, mais leur valeur

## Cinq mythes sur le data mining

---

elle-même.

Les revenus et les coûts sont calculés pour chaque client, les programmes de marketing et de GRC (gestion de la relation client) sont ensuite élaborés en fonction de la rentabilité de ces clients. Telle compagnie de télécommunication sud-américaine anticipe et intervient pour empêcher le départ de clients à valeur élevée en repérant les schémas qui conduisent à la perte de clientèle, s'appuyant pour cela sur l'analyse des mesures d'utilisation, d'achat et de qualité de services. Tel autre opérateur européen de téléphonie mobile a utilisé le data mining pour analyser le risque de churn (tendance du consommateur à changer de prestataire) afin de prendre des mesures proactives ciblées visant à identifier les clients et les segments ayant la plus forte propension au churn. Armé de cette information, l'opérateur a lancé des campagnes de marketing ciblées qui lui ont permis de réduire de 50 % le taux de churn sur les segments concernés. En parallèle, la compagnie a bénéficié d'un taux d'augmentation de 30 % au cours de la campagne marketing, améliorant du même coup la satisfaction des clients et par conséquent leur durée de fidélité. Mais le data mining peut aller bien plus loin qu'une simple analyse du comportement de la clientèle. Un des principaux opérateurs aux USA utilise le data mining pour améliorer les performances de son réseau, sans avoir à dépenser des millions de dollars pour actualiser son infrastructure réseau existante. Grâce à une surveillance constante des règles de performances et à une analyse permanente de l'historique de l'utilisation des composants et des lignes principales, complétées par des mesures régulières de l'activité de son réseau, cet opérateur peut veiller à ce que les appels soient acheminés en utilisant la capacité disponible. Ces efforts lui permettent d'enregistrer une baisse des défauts de service et des interruptions techniques et, parallèlement, une hausse du nombre d'appels réussis passés sur son réseau. Non seulement le degré de satisfaction des utilisateurs augmente, mais l'opérateur est aussi en mesure de fournir la qualité et la disponibilité de service exigée (et mesurée) par les instances de régulation.

### **Mythe n° 2 : le data mining ne serait pas encore viable pour des applications professionnelles**

Le data mining est une technologie viable et ses résultats professionnels sont hautement prisés. Ceux qui propagent le mythe en question sont en général les mêmes qui se trouvent contraints d'expliquer pourquoi ils n'utilisent toujours pas cet outil. Leur discours s'articule autour de deux arguments liés. D'après le premier : " Il est impossible d'utiliser efficacement de grandes bases de données ". Quant au second, il stipule que : " Le data mining ne peut pas s'effectuer dans le moteur de l'entrepôt de données (data warehouse) ". Ces deux constats ont certes été vrais à une certaine époque. Mais n'oublions pas qu'il a longtemps été vrai aussi que les avions ne pouvaient pas quitter le sol.

Répondons simultanément à ces deux arguments. Les bases de données sont aujourd'hui tellement vastes que les entreprises craignent que l'architecture informatique supplémentaire nécessaire aux projets de data mining n'entraîne d'énormes coûts et que le traitement des données exigé pour chaque projet ne soit trop long. C'est oublier que certaines des bases de données modernes utilisent désormais une technologie de traitement en parallèle, qui autorise l'extraction de données au sein de la base. L'extraction de données dans la base permet aux entreprises de supprimer les déplacements de données, d'exploiter les performances du traitement en parallèle, de réduire la redondance des données et d'annuler les coûts de création et de maintenance d'une base de données distincte et redondante dédiée au data mining. La combinaison de l'extraction des données dans la base de données et du traitement en parallèle donne une technologie de data mining parfaitement viable.

La vraie question ne concerne donc pas la viabilité de la technologie du data mining mais consisterait plutôt à se demander : " Comment le data mining peut-il m'aider à améliorer l'activité de mon entreprise ? " Dans le secteur des télécoms, le data mining est une pratique désormais

éprouvée qui a donné de vrais résultats avec un retour sur investissement rapide dans de nombreux secteurs, notamment la relation client, la gestion des campagnes de marketing, la mesure de la valeur client, la gestion financière ou l'optimisation de l'assurance de revenus et des performances réseau, pour n'en citer que quelques-uns.

### **Mythe n° 3 : le data mining exigerait une base de données distincte et dédiée**

À en croire les prestataires de data mining, il serait indispensable de disposer d'outils onéreux et dédiés (base de données, data mart ou serveur analytique) pour mettre en place cette technique. Ils invoquent pour cela la nécessité d'extraire les données sous un format propriétaire pour un traitement efficace. Or, non seulement ces data marts sont coûteux à l'achat et à la maintenance, mais ils exigent que les données soient extraites pour chaque projet distinct de data mining, ce qui représente une procédure coûteuse et fortement consommatrice de temps.

Heureusement, les progrès réalisés dans le domaine des bases de données n'exigent plus d'effectuer le data mining dans un data mart distinct. En réalité, pour une opération de data mining efficace, il faut pouvoir disposer d'un entrepôt de données à l'échelle de l'entreprise, ce qui, au regard du coût d'investissement total, se révèle considérablement moins cher que d'utiliser des data marts distincts.

Voici pourquoi. Au fur et à mesure que les sociétés mettent en place des projets de data mining dans l'ensemble de l'entreprise, le nombre d'utilisateurs exploitant les modèles de data mining ne cesse de croître, de même que les besoins d'accès à de larges infrastructures de données. Un entrepôt de données d'entreprise d'avant-garde ne se contente pas de stocker efficacement l'ensemble des données de l'entreprise et de rendre largement inutiles d'autres data marts ou d'autres entrepôts. Il constitue également les fondations idéales pour des projets de data mining, en fournissant un répertoire de données unique à l'échelle de l'entreprise, capable d'offrir une vision cohérente et actualisée de

## Cinq mythes sur le data mining

---

la clientèle. En outre, l'intégration d'extensions de data mining au sein de l'entrepôt de données permet à l'entreprise de réaliser deux types d'économies supplémentaires. Tout d'abord, il n'est plus nécessaire d'acheter et d'assurer la maintenance de matériel supplémentaire uniquement dédié au data mining. Ensuite, les entreprises voient se réduire la nécessité d'effectuer un va-et-vient de données depuis et vers l'entrepôt pour réaliser les projets de data mining, ce qui, nous l'avons vu, est une opération gourmande en travail et en ressources.

Par exemple, l'un des plus grands opérateurs de téléphonie mobile aux USA utilise un EDW (Enterprise Data Warehouse, *entrepôt de données d'entreprise*) centralisé pour fournir des informations à une large gamme d'applications commerciales, depuis l'attention à la clientèle jusqu'au marketing. Il n'en a pourtant pas toujours été ainsi. Cet opérateur, desservant de nombreux marchés régionaux, a été conduit à établir différents data marts régionaux. Obtenir une vue holistique des informations à l'échelle de l'entreprise était devenu pour le moins difficile. Les prises de décision souffraient de l'absence de données ou d'informations cohérentes, ce qui affectait le moral d'une équipe qui ne pouvait pas visualiser le fruit de son travail.

En transférant le tout vers un EDW centralisé, cet opérateur a pu constater des résultats immédiats. Les données étaient devenues cohérentes, les décisions pouvaient être prises en toute sécurité, avec l'assurance que les données reposaient sur des fondations saines. En termes de performances, le nouvel EDW délivrait des informations cohérentes presque 90 % plus vite qu'avec l'ancienne approche par data marts fragmentés. Lorsqu'il est nécessaire d'analyser de grandes quantités de données, de tels gains de performances font la différence entre être présent le premier sur un marché en se fiant à son seul instinct et s'y présenter en sachant que les décisions prises l'ont été sur la base d'informations tirées du comportement réel des clients.

### **Mythe n° 4 : il faudrait être polytechnicien pour faire du data mining**

Certains estiment que le data mining est une technologie si complexe qu'il faut au moins trois polytechniciens pour la maîtriser : un spécialiste en statistiques ou en méthodes quantitatives, un commercial fin connaisseur de la clientèle et un informaticien.

La vérité oblige à dire que bien des projets réussis ont été menés à bien sans l'aide d'un seul de ces polytechniciens. On peut citer l'exemple d'une société sud-américaine de télécommunications qui a réussi à repérer des changements dans le comportement de ses clients, ce qui lui a permis de conserver 98 % de ses clients à valeur élevée en pleine période de dérégulation. Ce succès est à mettre au crédit d'une équipe multidisciplinaire travaillant en collaboration.

Le data mining demande un effort de collaboration de la part d'un personnel qualifié dans trois domaines. Les commerciaux doivent guider le projet en créant une série de questions commerciales spécifiques, puis interpréter les résultats obtenus. Des concepteurs de modèles analytiques, formés aux techniques, aux statistiques et aux outils du data mining, doivent construire un modèle fiable. Les informaticiens, de leur côté, doivent apporter leurs compétences dans le traitement et la compréhension des données, ainsi que leur indispensable soutien technique.

### **Mythe n° 5 : le data mining serait réservé aux grandes entreprises disposant d'un large volume de données client.**

La réalité est que toute entreprise, quelle que soit sa taille, possédant des données qui reflètent avec précision son activité où sa clientèle peut s'en servir pour construire des modèles et en tirer des informations précieuses sur ses défis commerciaux majeurs. Le volume des données clients d'une société n'a jamais été en cause.

Par exemple, un opérateur de taille moyenne téléphonie mobile dans le secteur Asie/Pacifique se trouvait confronté à un problème de fraude. Il a fait appel à une base de données centralisée pour analyser

plus vite l'historique détaillé de ses appels et pour mieux comprendre le comportement de sa base client. La détection des abonnements et des utilisations frauduleux n'a ainsi pris que quelques jours au lieu de plusieurs mois et l'opérateur a pu constater une diminution des pertes dues à la fraude de 60 % par compte. À l'heure où les opérateurs subissent une pression croissante les poussant à augmenter leur nombre d'abonnés et à améliorer le revenu moyen par utilisateur, la possibilité d'agir positivement sur le bilan sans avoir à conquérir de nouveaux abonnés ni à augmenter l'ARPU a eu des conséquences bénéfiques notables sur les performances financières de l'opérateur en question. Bien qu'il ait dû acheter la base de données pour traiter spécifiquement son problème de fraude, l'opérateur en a rapidement tiré bien d'autres profits. Les économies réalisées grâce aux opérations informatiques et à l'intégration des systèmes ont permis de raccourcir les cycles de développement d'applications en exploitant les données provenant de l'entrepôt de données, et de réduire les coûts de 30 à 50 %. Une fois ces bénéfices constatés, des départements de plus en plus nombreux se sont mis à utiliser la base de données, permettant ainsi l'accès rapide aux informations commerciales stratégiques et de promptes prises de décisions dans des domaines tels que le churn, les activations quotidiennes et les flux de revenus. Les utilisateurs ont pu constater que des informations qui n'étaient auparavant accessibles qu'au bout de plusieurs semaines ou plusieurs mois étaient désormais accessibles de façon quotidienne.

### **Profitez-en**

En conclusion : le data mining n'est plus aujourd'hui ni lent, ni coûteux ni d'utilisation trop complexe. La technologie et le savoir-faire commercial existent et permettent de mettre en place un processus efficace et économe. Des compagnies de télécommunications de tailles diverses font partie des entreprises qui ont décidé de mettre les vieux mythes à l'épreuve et de prouver que le data mining est un processus essentiel pour pouvoir prospérer dans un univers professionnel dominé par une concurrence féroce et par l'obsession du client.

Pour de plus amples informations sur Teradata, veuillez vous rendre sur le site [www.teradata.com](http://www.teradata.com)

**Teradata**  
a division of  NCR