



CRM Analytique L'apport du Data Mining

Françoise Soulié Fogelman
francoise@kxen.com

2èmes Journées thématiques
Apprentissage Artificiel & Fouille de Données
jeudi 27 et vendredi 28 avril 2006
Université Paris 13, Institut Galilée



Le CRM analytique dans les entreprises

- **De plus en plus d'utilisateurs potentiels**
 - Utilisateurs métier ET PAS experts (BI, data mining)

- **Les entreprises veulent produire de plus en plus**
Typiquement par an
 - ◆ Des 1 000s de Tableaux de Bord,
 - ◆ Des 100s de modèles

- **Ce qui exige**
 - **Plus de gens doivent pouvoir produire ces TdB / modèles**
 - **Or il y a**
 - ◆ Beaucoup d'utilisateurs métier (plusieurs 1 000 000 s)
 - ◆ Peu d'experts (quelques 100 000 s)
 - **Mais il faut alors que la production de TdB / modèle soit « fool-proof » et automatisée**



Le CRM analytique dans les entreprises

- **Le déploiement des TdB est largement réalisé**
 - Grâce à des outils simples d'utilisation et adaptés au besoin
- **Le déploiement large du data mining nécessite une (r)évolution :**
 1. **Automatisation du codage de données & de la sélection de variables**
 2. **Utilisation de méthodes de sélection automatisée de modèle (SRM, régularisation ...)**
 3. **Utilisation de modèles simples (régression linéaire) pour expliquer facilement**
 4. **Utilisation de techniques garantissant la robustesse (ensemble test / cross validation)**

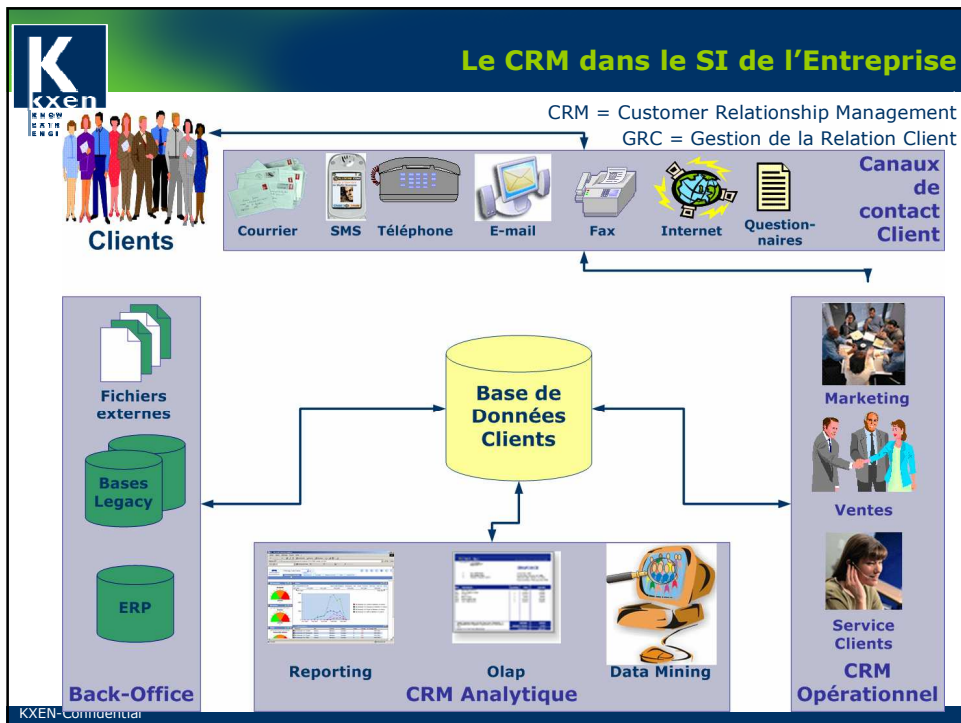
KXEN-Confidential



Agenda

- **Qu'est ce que le CRM analytique ?**
- **Les données**
- **L'apport des analyses prédictives**

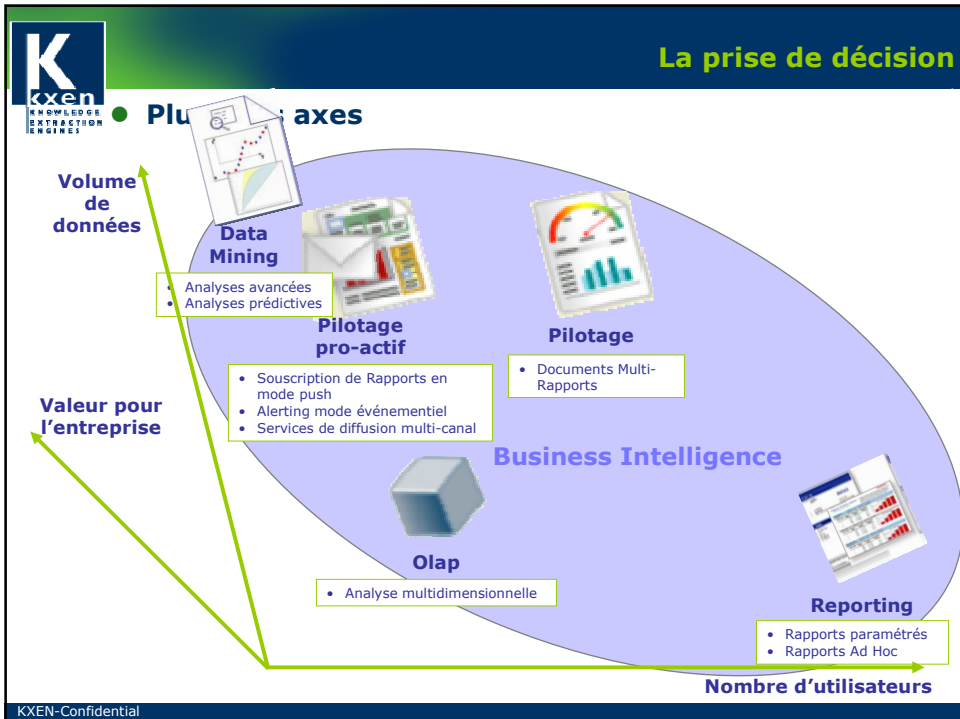
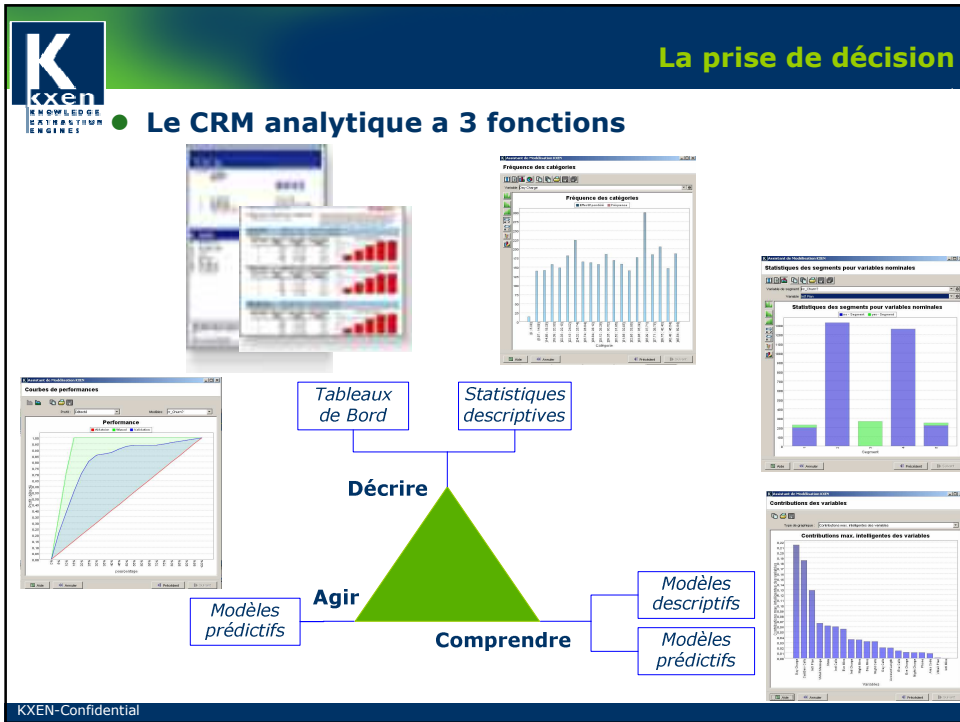
KXEN-Confidential



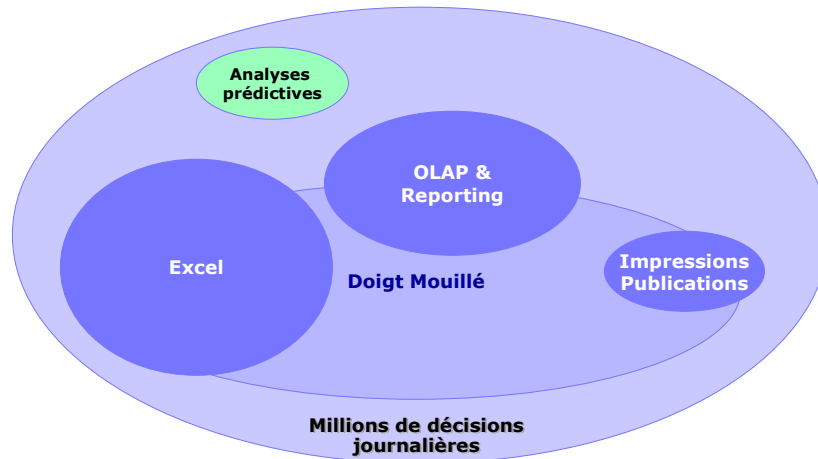
Le constat

- **Des sources de données nombreuses**
 - Le volume des données double chaque année (*Forrester*)...
 - Mais les sources sont hétérogènes, incohérentes
- **Des points de contact clients multiples**
 - sources d'informations clients riches ... mais hétérogènes
- **Des utilisateurs – et des besoins – nombreux**
 - Les utilisateurs souhaitent répondre par eux-mêmes à leurs questions, sans dépendre d'experts ni être obligés d'en devenir eux-mêmes.
http://www.decideo.fr/
- **Des décisions de plus en plus nombreuses**
 - Mais la qualité des décisions prises dépend des analyses menées
 - « Les profits sont les premiers à souffrir d'une prise de décision médiocre, de même que la réputation d'une société, son chiffre d'affaires, son service client et le maintien de ses marges bénéficiaires. »
 - « La qualité des décisions s'est améliorée ... principalement (du fait) d'analyses plus performantes, d'un accès plus rapide à des informations détaillées, et d'une expérience plus poussée »
Etude Teradata citée dans http://www.decideo.fr/

KXEN-Confidential



- **Un processus encore artisanal**



- **Qu'est ce que le CRM analytique ?**
- **Les données**
- **L'apport des analyses prédictives**

KXEN KNOWLEDGE EXTRACTION ENGINEER

Les données client

Les entreprises ont créé des datawarehouses client

- **En collectant l'ensemble des données disponibles**
 - ◆ En back-office
 - ◆ A travers l'ensemble des canaux de contact client
- **... mises en cohérence et intégrées dans le datawarehouse**

Bases de Production

- Fichiers
- Systèmes Legacy
- Fichiers externes

Points de Contact Clients

- Centre d'appel
- Téléphone
- Fax
- Courrier
- SMS / MMS
- e-mail
- Web
- Point de vente

Accès aux données → ODS → Manipulation des données → Data Warehouse

KXEN-Confidential

KXEN KNOWLEDGE EXTRACTION ENGINEER

Les données client

Les datawarehouses client

- **La taille (jusqu'à 100s TO) croit vite (X3 tous les 2 ans)**
- **Investissements de 100s M\$**
- **Millions de clients & milliers de variables**

Taille de la Base (TO)

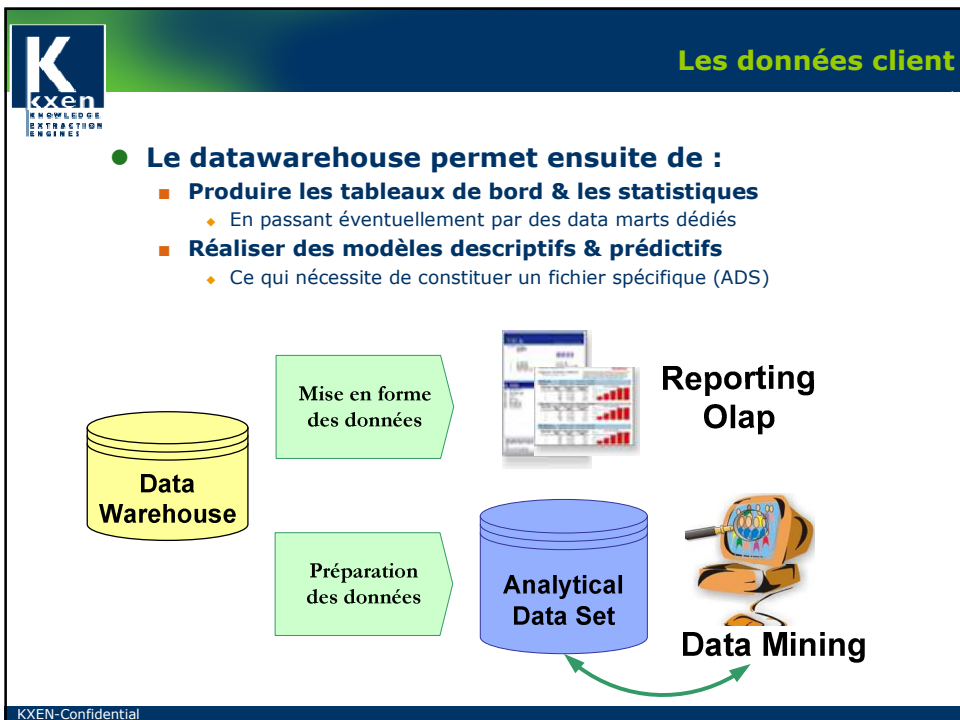
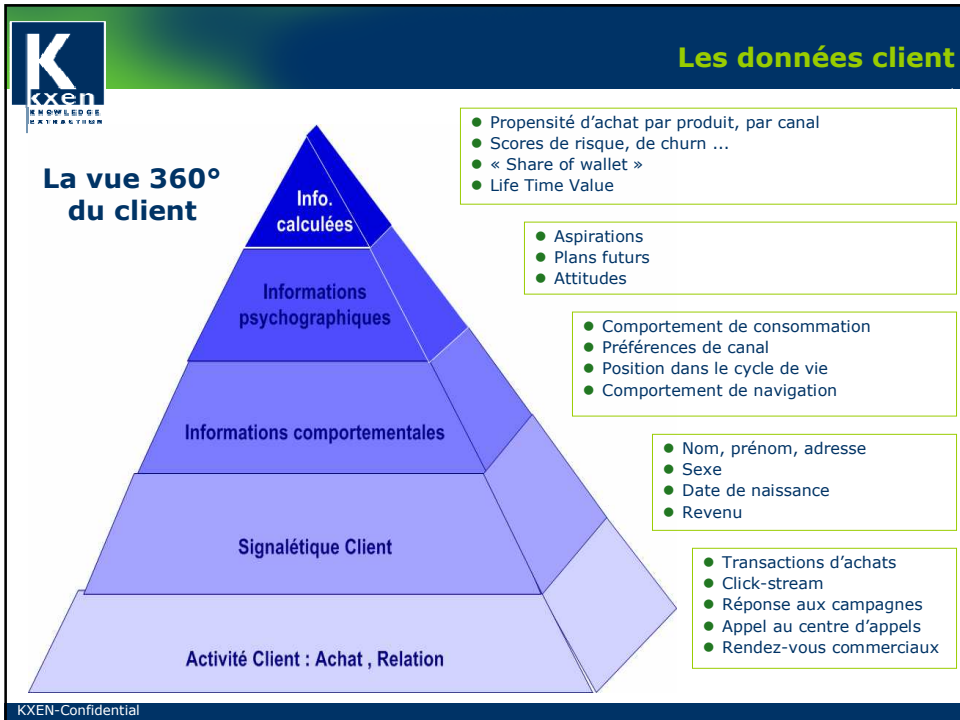
Top Ten	2005	2003	2001
1	100	30	10
2	90	25	5
3	50	25	5
4	25	15	5
5	25	10	5
6	25	10	5
7	20	10	5
8	20	10	5
9	20	10	5
10	20	10	5

Nb de lignes, enregistrements ou objets (Millions)

Top Ten	2005	2003	2001
1	2800000	500000	100000
2	1800000	200000	50000
3	500000	100000	50000
4	500000	100000	50000
5	400000	100000	50000
6	200000	100000	50000
7	200000	100000	50000
8	200000	100000	50000
9	200000	100000	50000
10	200000	100000	50000

D'après <http://www.wintercorp.com>

KXEN-Confidential



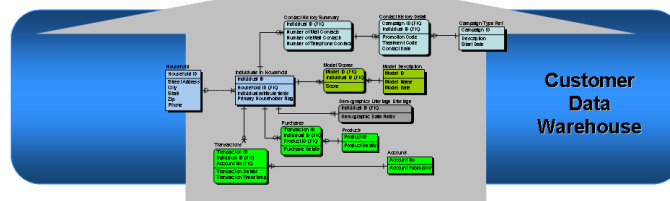
Pour produire le modèle prédictif il faut :

- **Un Analytical Data Set ou Customer Analytic Record**
 - Récupérer des données dans des sources multiples, disponibles dans le SI ou directement dans le datawarehouse
 - Intégrer ces données dans l'ADS (un "fichier plat")

IPAD	COMPID	NAME	VALUE_360	BEHAV_360	DEMO_360	SCORE_360	CONTACT_HIST
2347387474	4797978688	Quintavo	2	5	8	37.99	78.18	49.8	14
7878978978	2438970274	Susan	18
9870908	879878	Andre	12
...

virtual flat record

Customer Analytic Record

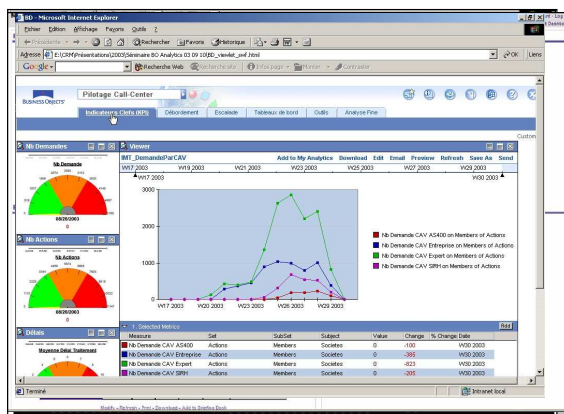


Constituer un datawarehouse clients exige de gérer l'ensemble des données client à travers tout leur cycle de vie

- **Capturer les données clients**
- **Nettoyer les données**
 - Normalisation d'adresses, mise en cohérence
 - Déduplication
- **Enrichir les données**
 - Données externes
- **Constituer la Base Clients (DWH)**
 - Données normalisées
- **Distribuer les données aux utilisateurs en fonction de :**
 - Leurs besoins
 - Leurs droits
- **... le tout en respectant les contraintes**
 - Qualité
 - Fréquence de rafraîchissement / temps de chargement
 - Synchronisation
 - Disponibilité temps réel si nécessaire

- **Qu'est ce que le CRM analytique ?**
- **Les données**
- **L'apport des analyses prédictives**

- **Ils mettent en forme les données historisées ...**
 - **Indicateurs & axes d'analyse structurés en hiérarchies métier**
 - **Exploration des données – drill down**
 - **Alertes adaptées aux besoins métier**
 - **Support pour la Direction, les utilisateurs métier**
- **... du passé**



- **Intégrer toutes les données (historiques)**
- **Définir les indicateurs & les axes d'analyse**
 - Fiables, explicites, faciles et rapides à obtenir
 - Problème du choix de la granularité
- **Définir les rôles (profils utilisateurs)**
 - Les droits d'accès
 - La personnalisation
- **Mettre en œuvre les rapports**
- **Donner accès aux utilisateurs / rôle / canal**
- **... le tout en respectant les contraintes**
 - Qualité
 - Fréquence de rafraîchissement / temps de chargement
 - Facilité d'emploi pour les utilisateurs
 - Adéquation au métier des utilisateurs
 - Nombre adéquat de rapports («trop d'information tue l'information»)
 - L'utilisateur peut se passer « d'expert », il peut trouver l'information utile avec un temps d'apprentissage rapide

Reporting

Analyse

Analyse prédictive

● Le reporting

- N'est « que » un mode de présentation des données
- Les données sont normalisées
- Permet de constater a posteriori ce qui s'est passé
- L'« intelligence » est fournie par l'utilisateur qui peut
 - ◆ Analyser les informations présentées dans les tableaux de bord
 - ◆ En tirer les conclusions pour définir les meilleures actions
- Permet d'agir en réaction au passé, à condition d'analyser correctement les rapports
- Apporte une « valeur » à l'entreprise difficile à mesurer

● L'analyse prédictive

- Exploite les données du passé pour fournir des prévisions
- Les données sont « à plat »
- Permet de prévoir ce qui va se passer dans le futur
- L'« intelligence » est fournie par l'outil data mining
 - ◆ Sélection des variables significatives & scores prévisionnels
 - ◆ Listes pour optimiser les actions et mesures des performances attendues
- Les résultats sont directement exploitables dans des actions CRM en mode pro-actif
- C'est un facteur majeur d'augmentation des revenus
 - ◆ Il est facile de construire un business case

Un exemple dans les telecom (moyennes en France)

	Cas 1	Cas 2
Base Clients	5 000 000	10 000 000
ARPU / mois	45 €	35 €
Coût d'acquisition	200 €	150 €
CA sur 1 an (M €)	2 700	4 200

ARPU : Average Revenue per User

Le churn

- Un client part à la concurrence

Questions

1. Evaluer et suivre le coût du churn
 - Comment ?
2. Réduire le churn
 - Qui doit-on retenir ?
 - Pourquoi ?
 - Comment ?



1. Evaluer et suivre le coût du churn

- **Méthode 1**
 - ◆ On évalue simplement la perte de CA an
- **Méthode 2**
 - ◆ On ajoute aussi les coûts d'acquisition (dépensés pour rien)
 - ◆ (Mais on a déjà commencé à émettre des factures mensuelles ...)

	Cas 1	Cas 2
Base Clients	5 000 000	10 000 000
ARPU / mois	45 €	35 €
Coût d'acquisition	200 €	150 €
CA sur 1 an (M €)	2 700	4 200
Taux churn / mois	2,00%	2,50%
Coût du churn (1) (M €)	648	1 260
Coût du churn (2) (M €)	668	1 298

- **On peut produire un TdB de suivi du churn selon plusieurs axes**
 - Type de forfait, montant de la facture mensuelle, géographie, temps ...

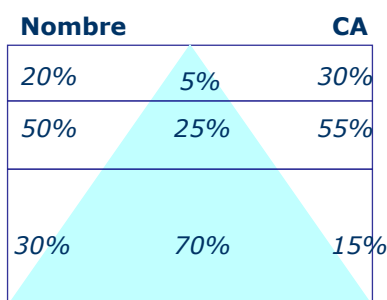
KXEN-Confidential



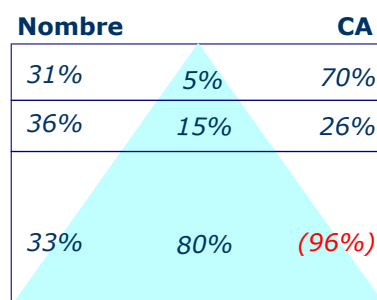
2. Réduire le churn

- **Tous les clients ne sont pas semblables ...**

Distribution ...

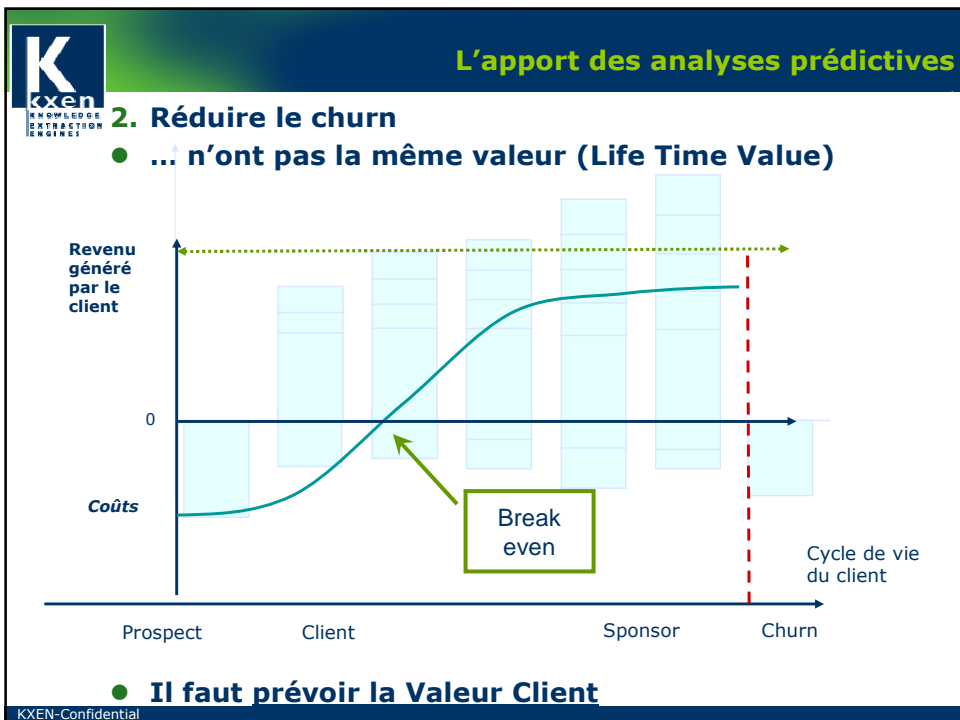
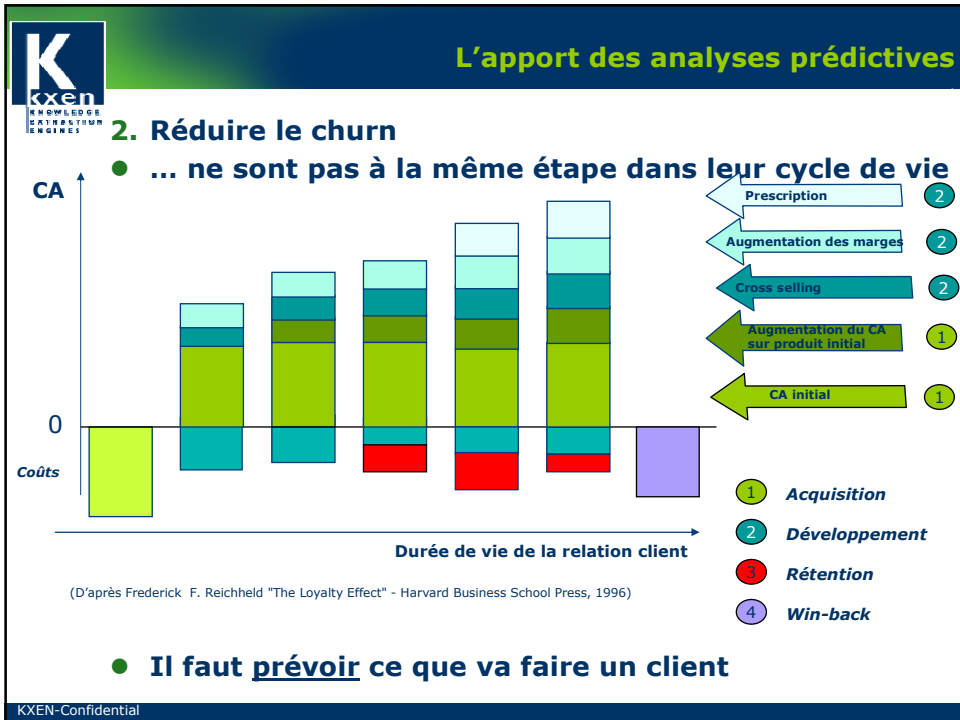


Banque ...



- **Il faut segmenter les clients**

KXEN-Confidential



K
KXEN
KNOWLEDGE
EXTRATION
ENGINEER

L'apport des analyses prédictives

2. Réduire le churn

- ... ne doivent pas être traités de la même façon www.1to1.com

Valeur client

- Il faut **personnaliser les actions**

KXEN-Confidential

K
KXEN
KNOWLEDGE
EXTRATION
ENGINEER

L'apport des analyses prédictives

2. Réduire le churn

- Prévoir ce que va faire un client
 - ◆ Pourquoi
 - ◆ Produire des listes de ciblage

Assistant de Modélisation KXEN

Contributions des variables

Type de graphique: Contributions max. intelligentes des variables

Contributions max. intelligentes des variables

Variables

Assistant de Modélisation KXEN

Courbes de performances

Protoc: Testé

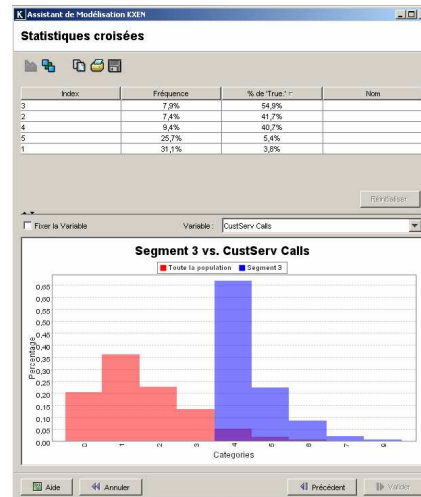
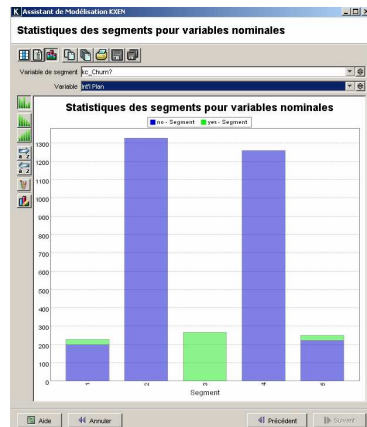
Modèle: P_Churn?

Performance

KXEN-Confidential

2. Réduire le churn

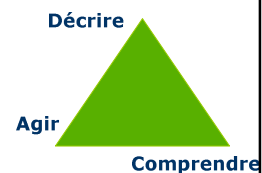
- **Segmenter les clients**
 - ◆ Produire les segments
 - ◆ Caractériser les segments



KXEN-Confidential

● Qu'apportent les analyses ?

- **Connaissance client**
 - ◆ Comprendre ce qui différencie les clients
 - ◆ Détecter les leviers d'actions
- **Fournir les informations nécessaires pour les actions**
 - ◆ Gagner en efficacité
 - ◆ Exemple : campagne ciblée / non ciblée



	Campagne non ciblée	Campagne ciblée
Base Clients	5 000 000	5 000 000
Nb clients ciblés / an	2 500 000 €	2 500 000 €
Taux de réponse	4%	8%
Revenu généré / réponse	30 €	30 €
Retour des campagnes	3 000 000 €	6 000 000 €
Apport du ciblage		3 000 000 €

- On peut aussi réduire la taille de la cible (et donc les coûts) tout en maintenant le nombre de réponses

KXEN-Confidential

K
KXEN
KNOWLEDGE
EXTRATION
ENGINEER

Wish list

- **Manipuler des données**
 - Volumes croissants
 - Sources hétérogènes
- **Produire des modèles**
 - Vite, avec précision
 - En temps réel
- **S'adapter aux besoins d'utilisateurs multiples**
 - Les utilisateurs métier veulent pouvoir faire eux-mêmes
 - ... mais ne veulent pas devenir des experts techniques
- **Intégrer dans les Systèmes d'Information**
 - Lire tout format de données
 - Ecrire vers tout format
- **Produire de la valeur**
 - Modèles précis, utilisables
 - Productivité

Eviter de dupliquer les données

Entraîner les modèles ... vite

Appliquer les modèles ... vite

Accessible aux utilisateurs métier

Conforme aux standards Data mining

Retour Sur Investissement

KXEN-Confidential

K
KXEN
KNOWLEDGE
EXTRATION
ENGINEER

Approche Data mining

Approche traditionnelle

Approche industrielle

- **Productivité**
 - Automatisation
 - ◆ Codage, pas de recherche d'algorithme, test intégré, debriefing intégré
- **Efficacité**
 - Robustesse : Vapnik
- **Méthodologie performante**
 - KDMAIC

KXEN-Confidential

KXEN KNOWLEDGE EXTRACTION ENGINEER

Approche KXEN

Vapnik's SRM

- **KXEN is based upon Structured Risk Minimization**
... using classes

$$\Phi_n = \{f(\cdot) / f: \text{polynôme de degré } n\}$$

$$\Phi_c = \{f_a(\cdot) / f_a(x) = a.x, \|a\| \leq c\}$$

KXEN-Confidential

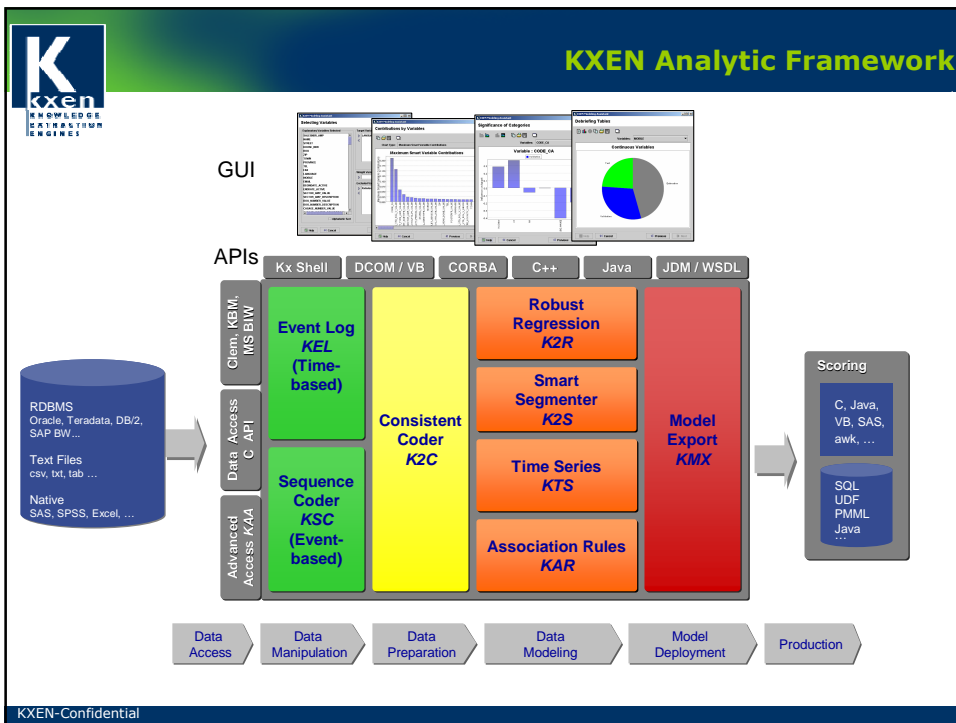
KXEN KNOWLEDGE EXTRACTION ENGINEER

Approche KXEN

Codage non-linéaire de chaque variable
+ **régression polynômiale (linéaire en général)**
... **contrôlé par SRM**

(compromis fit/robustness)

KXEN-Confidential




KXEN Analytic Framework

Exemples

- **Modèles réalisés en 2 jours**
- **Coûts opérationnels réduits de 50% et time-to-model de 90%**
- **Score 75 M foyers dans Teradata en 30 minutes**
- **20 modèles prédictifs produits & déployés par 2 personnes en 1 mois**
- **Réduit le temps de préparation des données de 70% de l'effort total à (presque) rien**
- **Maintenance automatique de centaines de modèles / an**
- **Gains de productivité attendus 10x**
- **1680 modèles en production / an**
- **"I built 377 models with 100K rows each in a couple hours on only a PC"**
- **10 modèles en 3 jours**

Logos of partner companies: Sears, BARCLAYS, vodafone, COX COMMUNICATIONS, belgacom.

KXEN-Confidential



Exemple

Vodafone

- **Base Teradata**
 - 2500 variables
- **Environnement analytique**
- **700 modèles / an**
- **Production de très nombreux modèles**
- **Résultats d'analyses disponibles en temps réel**
- **Performance accrue**

So, is this all that is needed? No

- Ideal analytical environment also includes:
 - Automated, over-night, creation of new analytic variables
 - Tighter integration with campaign management systems, profiling and direct link into actions
 - Measurement system to show effectiveness of all scores over time
 - This will be important when there are hundreds of models

But, what we are going to build is a HUGE improvement versus what Vodafone has now.

experience the POSSIBILITIES 2005


experienced analysts

experience the POSSIBILITIES 2005

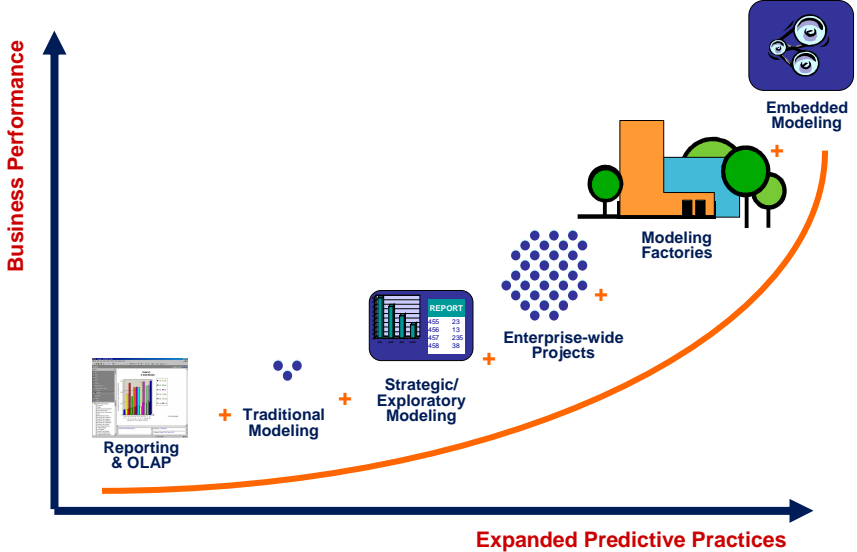
This means trying to create 716 models per year... 5

<http://www.teradata.com/teradata-partners/conf2005/>

KXEN-Confidential



L'usine à modèles



Business Performance

Expanded Predictive Practices

KXEN-Confidential



Conclusion

- The analytical functions necessary to support mining of nonlinear data sets must become integrated into the very structure of other software tools, **similar to that of arithmetic operations.**
- Data mining is not really an "end" per se , but a means to an end.
- These "means" will become progressively submerged in the infrastructure of the products they serve until **they are as natural to use as standard arithmetic and graphical techniques.**

Robert A. Nisbet

http://www.dmreview.com/editorial/newsletter_article.cfm?articleId=1050627

KXEN-Confidential



QUESTIONS ?

KXEN-Confidential