

# Le Data Mining

## Techniques pour exploiter l'information



M.E.S Gestion Sàrl  
Rue Louis-Favre 34  
2002 Neuchâtel  
☎ +41 32 729 35 00  
✉ +41 32 729 35 01  
contact@mesgestion.ch  
www.mesgestion.ch

Auteur : Dan Noël

Date : 24.04.2009



# Agenda de la présentation du 26.03.2009

- Concept de Data Mining ou qu'est-ce que le Data Mining
- Déroulement d'un projet de Data Mining
- Place du Data Mining dans le système d'information
- Principales applications Marketing et techniques de Data Mining
- Avantages et résultats observés
- Freins et blocages au développement du Data Mining
- Conseil « quantitatif »: réalité ou vision de l'esprit ?
- Questions & réponses

# Le défi: La prise de décision intégrant l'ensemble de l'information disponible

„In our view, the winners will be those companies that develop and use information to seize market share from their less talented competitors.”

Analyst,  
Morgan Stanley Dean Witter

# Is Big Brother watching you?

Vous et moi laissons toujours plus de traces permettant de caractériser nos comportements d'achats, nos habitudes, nos préférences,...

- **Cartes débit/crédit**
- **Cartes des grands distributeurs (Cumulus, SuperCard,...)**
- **Achats par correspondance**
- **Internet**
- **Les sources officielles (impôts,...)**
- **Les clubs**
- **Les études de marché / sondages**
- ...

# Augmentation de la capacité d'analyse grâce aux développements

- Création et stockage de grandes quantités de données par l'automatisation des procédés
- Disponibilité d'une forte capacité de calcul pour un investissement modéré
- Coûts décroissants du stockage des informations
- Nouvelles méthodes techniques et scientifiques

# De plus en plus de données disponibles...

- Explosion des données clientèle disponibles (intra et extra entreprise)
  - **Socio-démographique**
  - **Profil clients (loisirs, habitudes,...)**
  - **Comportement « d'achat » (volume, type, préférence canaux,...)**
  - **Critères d'achat (études clients,...)**
- « Trop de données tue l'information!!! »



# La qualité des informations dépend des sources

- Connaissances des experts (expériences, formation)
- Données opérationnelles sur les clients et leurs comportements transactions)
- Sources externes (analyse du marché, presse Internet, fournisseurs externes)
- Les données doivent être traitées, analysées, évaluées et transformées en connaissances utiles à la prise de décision

# Réponse: le Data Mining

## ● Le Data Mining:

- Ensemble de méthodes et de techniques d'analyse de données et d'extraction d'informations structurées en vue d'aider à la prise de décision.

## ● Le Data Mining descriptif:

- Mettre en évidence des informations présentes mais noyées par le volume de données.

## ● Le Data Mining prédictif:

- Extrapoler des nouvelles informations à partir de données existantes
- Anticiper les besoins de la clientèle et adapter la stratégie de l'entreprise aux attentes du marché



# Concept de Data Mining

- La richesse d'une entreprise = ses clients
- Objectifs principaux des entreprises
  - Augmenter la rentabilité et la fidélité des clients
  - En maîtrisant les risques
  - Et vendre au travers des bons canaux, au bon moment, les bon produits...



# Utilité du Data Mining au niveau commercial

- Mieux connaître ses clients:
  - Pour mieux les servir
  - Pour augmenter leur degré de satisfaction
  - Et par conséquent leur fidélité
  - Et donc leur rentabilité
- Particulièrement pertinent dans le tertiaire:
  - «Produits» quasiment substituables
  - L'élément prix n'est pas forcément déterminant
  - La relation client et le service font la différence



# Exercice: la valeur d'un client

- Nombre de client début de la période: 100'000
- Taux de rétention: 80%
- Transaction/client: CHF 200.-/an
- Coûts variables: 70%
- Facteur d'actualisation: 10%

● Exercice:

**Calculer la valeur client long-terme (CLV) et la valeur du portefeuille actuel.**

# Exercice: la valeur d'un client

## Réponse

	1ère année	2ème année	3ème année
<i>Nombre de clients</i>	100'000	80'000	64'000
<i>Taux de rétention</i>	80%	80%	80%
<i>Transaction / client</i>	200	200	200
<b>Chiffre d'affaires</b>	<b>20'000'000</b>	<b>16'000'000</b>	<b>12'800'000</b>
<i>Coûts variables %</i>	70%	70%	70%
<i>Coûts variables \$</i>	14'000'000	11'200'000	8'960'000
<b>Profit brut</b>	<b>6'000'000</b>	<b>4'800'000</b>	<b>3'840'000</b>
<b>Facteur d'actualisation (10%)</b>	1	1.1	1.21
<b>Profit net</b>	<b>6'000'000</b>	<b>4'363'636</b>	<b>3'173'554</b>
<b>Profit / client</b>	<b>60.00</b>	<b>43.64</b>	<b>31.74</b>

**Valeur client long-terme**

**135.4**

**Valeur du portefeuille**

**13'537'190**

# Les opportunités pour augmenter la valeur du profit (1/2)

- Réduire les coûts variables
- Augmenter le revenu par client
- Prolonger une relation profitable avec le client
- Acquérir des clients « profitables »



# Les opportunités pour augmenter la valeur du profit (2/2)

Retenir plus de clients



- Qui partira?
- Quand partira-t-il?
- Qui veut-on garder?

Plus de profit par client



- Qui est intéressé par des produits supplémentaires?
- Qui profite combien de fois de nos services?
- Qui achète les produits les plus profitables?

Plus de clients



- Quels clients veut-on acquérir?
- Qui sera profitable?

# Les solutions à ces questions

- Pour répondre à ces questions = analyse prédictive
- Prédiction précise = expériences pertinentes
- Un modèle est une expression mathématique d'évènements antérieurs
- Exemple: la probabilité d'achat d'un certain bien ou service augmente avec l'âge (probabilité d'achat =  $1.352 \cdot \text{âge}$  \*)

# Augmentation valeur client long-terme

	1ère année	2ème année	3ème année
<i>Nombre de clients</i>	100'000	85'000	72'250
<i>Taux de rétention (+6.25%)</i>	85%	85%	85%
<i>Transaction / client (+5%)</i>	210	210	210
<b>Chiffre d'affaires</b>	<b>21'000'000</b>	<b>17'850'000</b>	<b>15'172'500</b>
<i>Coûts variables % (-7%)</i>	65%	65%	65%
<i>Coûts variables \$</i>	13'671'000	11'620'350	9'877'298
<b>Profit brut</b>	<b>7'329'000</b>	<b>6'229'650</b>	<b>5'295'203</b>
<b>Facteur d'actualisation (10%)</b>	1	1.1	1.21
<b>Profit net</b>	<b>7'329'000</b>	<b>5'663'318</b>	<b>4'376'200</b>
<b>Profit / client</b>	<b>73.29</b>	<b>56.63</b>	<b>43.76</b>

**Valeur client long-terme**  
**Valeur du portefeuille**

**174** + CHF 38.4  
**17'368'519** + CHF 38'840'000



# Rentabilité des activités analytiques

- Augmentation de la valeur du portefeuille = 38'840'000
- Augmenter le revenu par client de CHF 10.-/an
  - Réduire les coûts variables de CHF 3.50 /client
  - Actions de « cross ou up selling)
  - Meilleure rentabilité des actions commerciales
- Conserver les clients rentables 2.6 semaines de +
- Possible grâce à une meilleure connaissance détaillée du client

# Quelques exemples d'utilisation du Data Mining dans une banque

- **Marketing** (gestion de la relation client, ciblage actions commerciales, cross-upselling, analyse des ventes, segmentation ...)
- **Risk-Management** (détection de fraudes, scoring crédit, détection blanchiment d'argent,...)
- **Product Management** (tarification produits, politique de packaging)
- ...



# Data Mining n'a pas réponse à tout

- Le Data Mining n'est pas une boule de cristal:
  - Il ne devine pas le profil des clients à cibler, ceux sur le départ,...
  - Il extrapole à partir des données disponibles
- Garbage In = Garbage Out



# Data Mining ≠ statistiques descriptives

- Le Data Mining permet l'analyse de phénomènes beaucoup plus complexes que les statistiques descriptives
- Il traite de combinaisons sous-jacentes et/ou cachées
- Il permet d'aller au delà du tâtonnement des statistiques descriptives conditionnée par la pré connaissance ou les stéréotypes de « l'intelligence de marché ».
- Avec lui, on passe de l'analyse confirmatoire à la véritable expérience exploratoire!
- Le Data Mining fait appel à des méthodes d'analyse bien plus complexes et qui peuvent parfois sembler « opaques »

# Définition du Data Mining

- Un ensemble de méthodes pour l'analyse et la gestion des données
- Utilisé pour l'extraction de connaissances à partir de grandes bases de données
- Permet de repérer les informations essentielles
- Représente les informations par des modèles compacts
- Met en correspondance les données avec la réalité

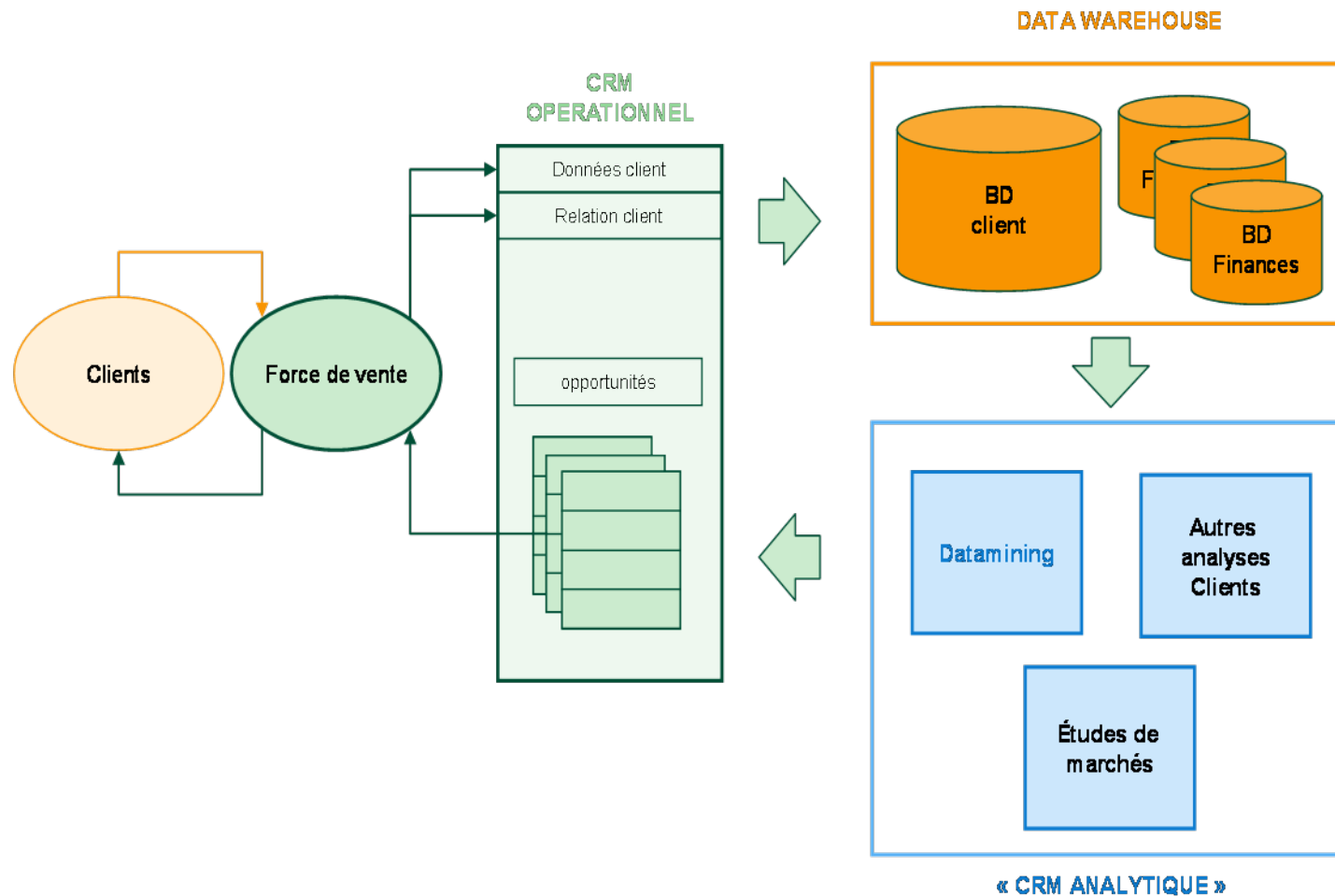
# L'historique des méthodes d'analyse

<b>Rapports d'analyse statiques (1980)</b>	<b>Rapports d'analyse dynamique (1990)</b>	<b>Data Mining (2000)</b>
Qui sont les clients dégagant le plus de chiffre d'affaires?	Parmi nos clients, quel ont été les 5% les plus actifs au cours des 5 dernières années?	Qui seront pour les 5 prochaines années, nos clients les plus rentables?
Déterministe, déductif, rétrospectif, bien défini, pas de modélisation. unidimensionnel	Déterministe, déductif, rétrospectif, bien défini, pas de modélisation. Multi-dimensionnel	Probabiliste, multi-dimensionnel, exploration, modélisation statistique
Bases de données relationnelles, SQL, rapports et listings	Data Warehouse, OLAP Cubes, rapports dynamiques, EIS	Statistiques, BI, classification, modélisation

# Les méthodes du Data Mining

Objectifs	Concepts	Techniques
Analyse exploratoire	Visualisation et transformation des données	Analyse uni et multi-dimensionnel + statistiques descriptives
Déduction de caractéristiques, règles et des tendances à partir des données existantes	Apprentissage supervisé	Régression linéaire Régression logistique Réseau de neurones Arbres de décision
Découverte de structures inconnues / sous-jacentes	Apprentissage non-supervisé	Clustering ACP / AFC Carte de Kohonen Basket-Analysis

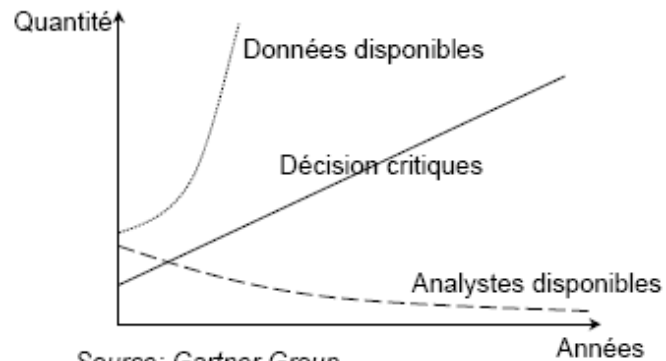
# Intégration du Data Mining dans le système d'information de l'entreprise



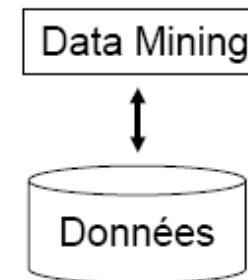


# Le Data Mining révèle les opportunités des données disponibles

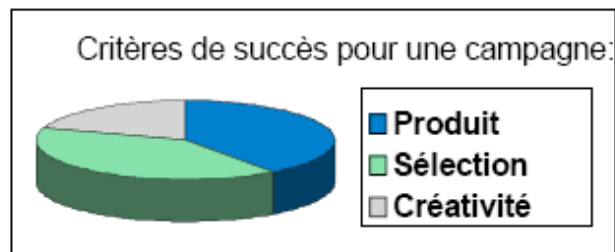
La différence des tendances...



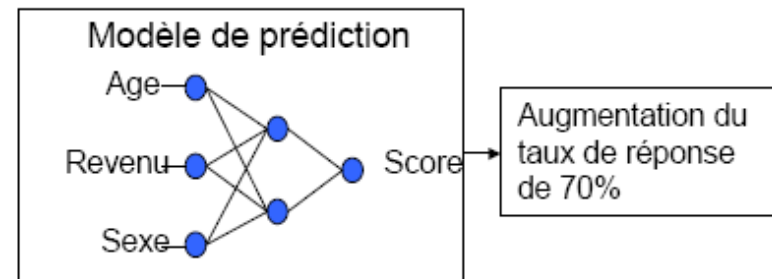
...nécessite une analyse rationnelle et efficace



Les connaissances ...

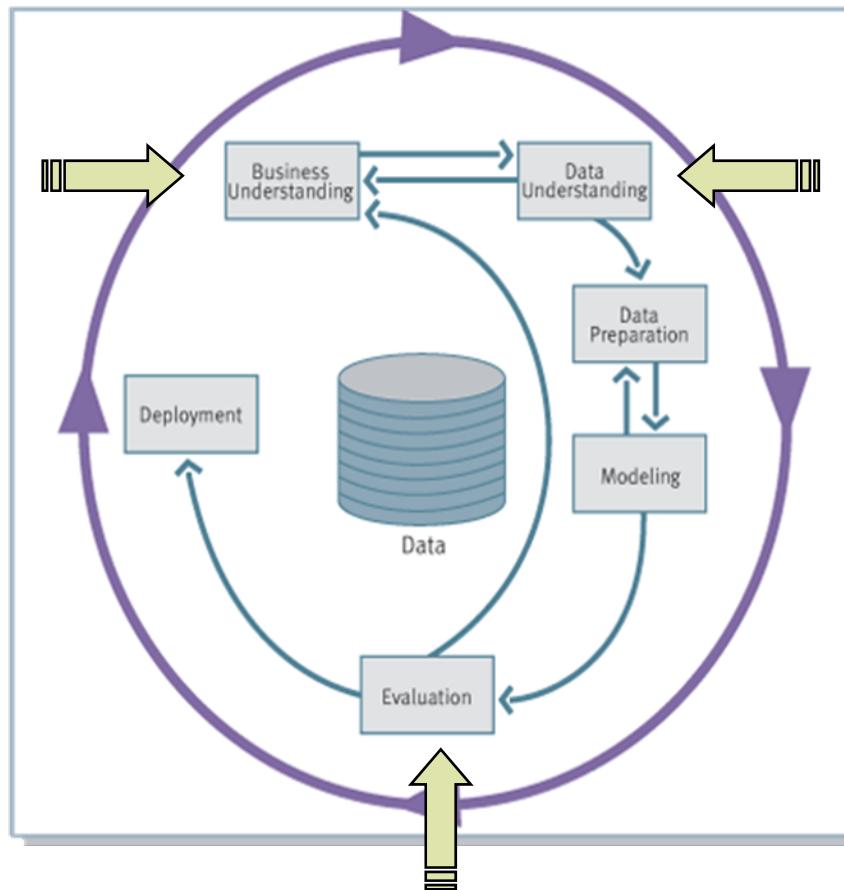


...augmentent le revenu



# Le Data Mining sous forme de processus

## Le modèle CRISP-DM



Principales interactions  
avec le business

Source: CRISP-DM 1.0, Step-by-step Data Mining Guide, SPSS Publication

# Les principales étapes d'un projet de DM

- Poser le problème et définir les objectifs
- La recherche des données
- Explorer et visualiser les données
- Sélection des données / réduction des dimensions
- Nettoyage des données
- Les actions sur les variables
- La recherche du modèle
- Évaluation des résultats
- Déploiement des résultats et suivi

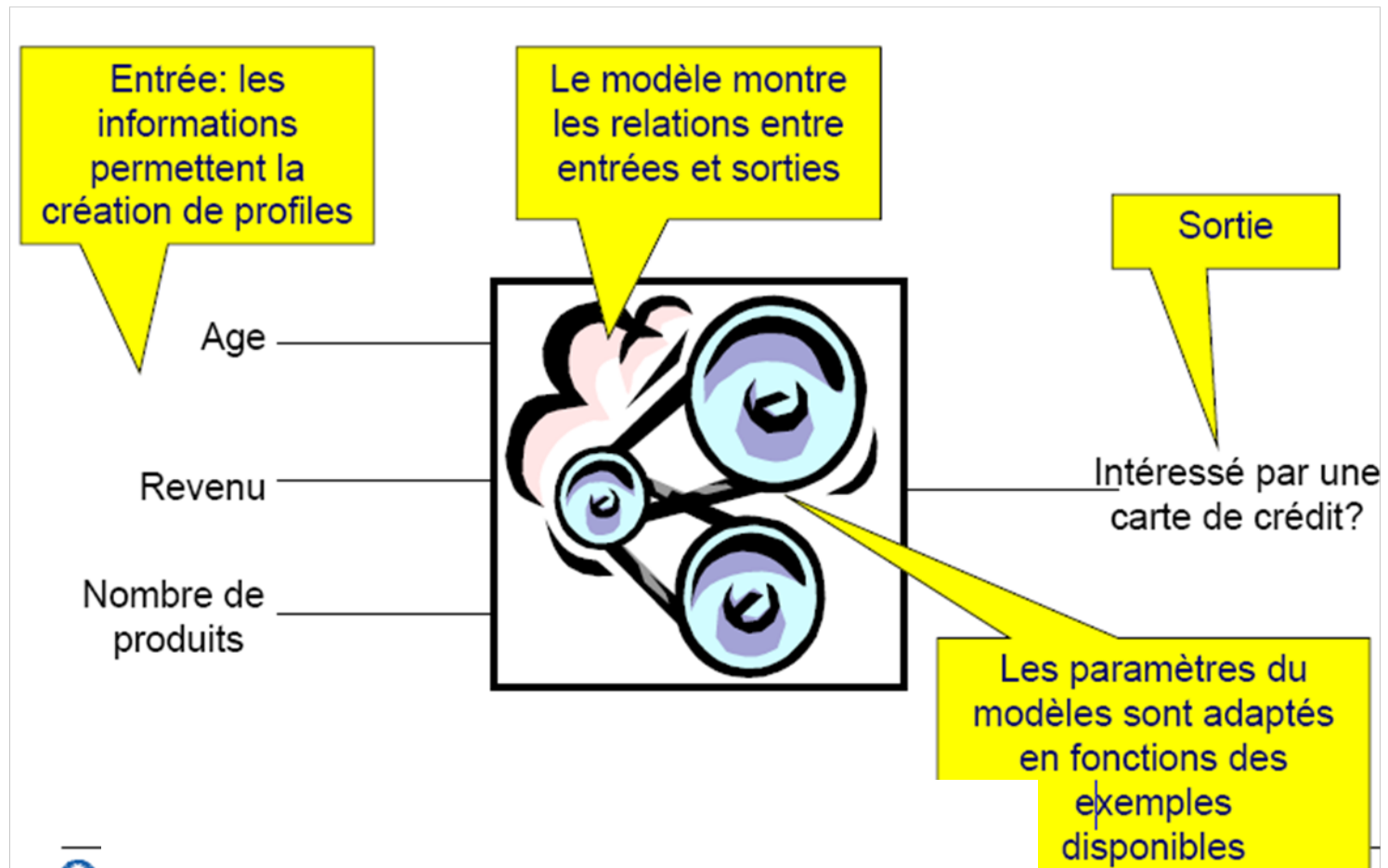
# Détail d'un projet de DM

## Selon le modèle CRISP-DM

Business Understanding	Data Understanding	Data Preparation	Modeling	Evaluation	Deployment
<p><b>Determine Business Objectives</b> Background Business Objectives Business Success Criteria</p> <p><b>Assess Situation</b> Inventory of Resources Requirements, Assumptions, and Constraints Risks and Contingencies Terminology Costs and Benefits</p> <p><b>Determine Data Mining Goals</b> Data Mining Goals Data Mining Success Criteria</p> <p><b>Produce Project Plan</b> Project Plan Initial Assessment of Tools and Techniques</p>	<p><b>Collect Initial Data</b> Initial Data Collection Report</p> <p><b>Describe Data</b> Data Description Report</p> <p><b>Explore Data</b> Data Exploration Report</p> <p><b>Verify Data Quality</b> Data Quality Report</p>	<p><b>Select Data</b> Rationale for Inclusion/ Exclusion</p> <p><b>Clean Data</b> Data Cleaning Report</p> <p><b>Construct Data</b> Derived Attributes Generated Records</p> <p><b>Integrate Data</b> Merged Data</p> <p><b>Format Data</b> Reformatted Data</p> <p>Dataset Dataset Description</p>	<p><b>Select Modeling Techniques</b> Modeling Technique Modeling Assumptions</p> <p><b>Generate Test Design</b> Test Design</p> <p><b>Build Model</b> Parameter Settings Models Model Descriptions</p> <p><b>Assess Model</b> Model Assessment Revised Parameter Settings</p>	<p><b>Evaluate Results</b> Assessment of Data Mining Results w.r.t. Business Success Criteria Approved Models</p> <p><b>Review Process</b> Review of Process</p> <p><b>Determine Next Steps</b> List of Possible Actions Decision</p>	<p><b>Plan Deployment</b> Deployment Plan</p> <p><b>Plan Monitoring and Maintenance</b> Monitoring and Maintenance Plan</p> <p><b>Produce Final Report</b> Final Report Final Presentation</p> <p><b>Review Project</b> Experience Documentation</p>

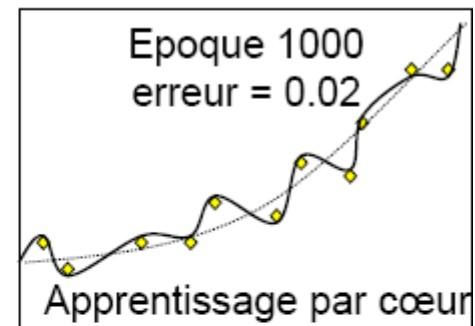
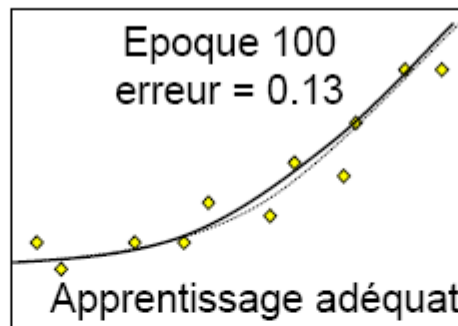
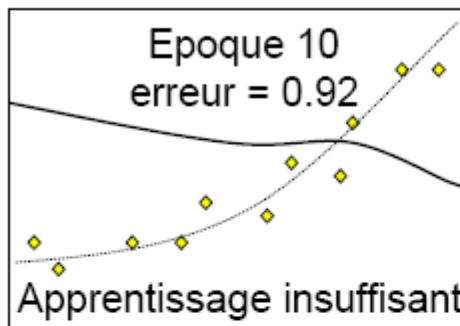
Source: CRISP-DM 1.0, Step-by-step Data Mining Guide, SPSS Publication

# Modélisation: la base du Data Mining

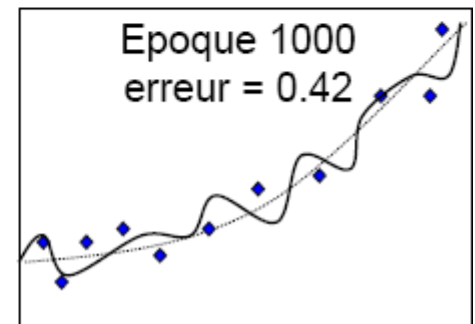
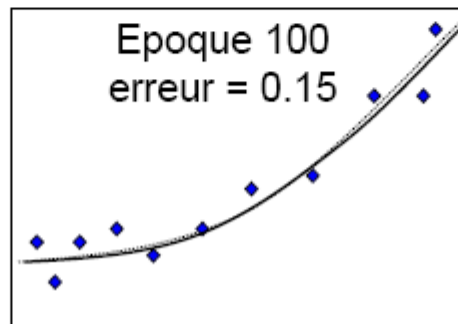
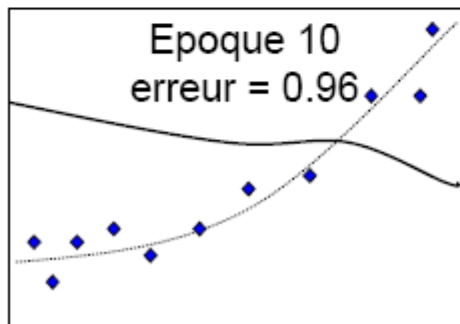


# Le phénomène de l'apprentissage

## Ensemble d'apprentissage

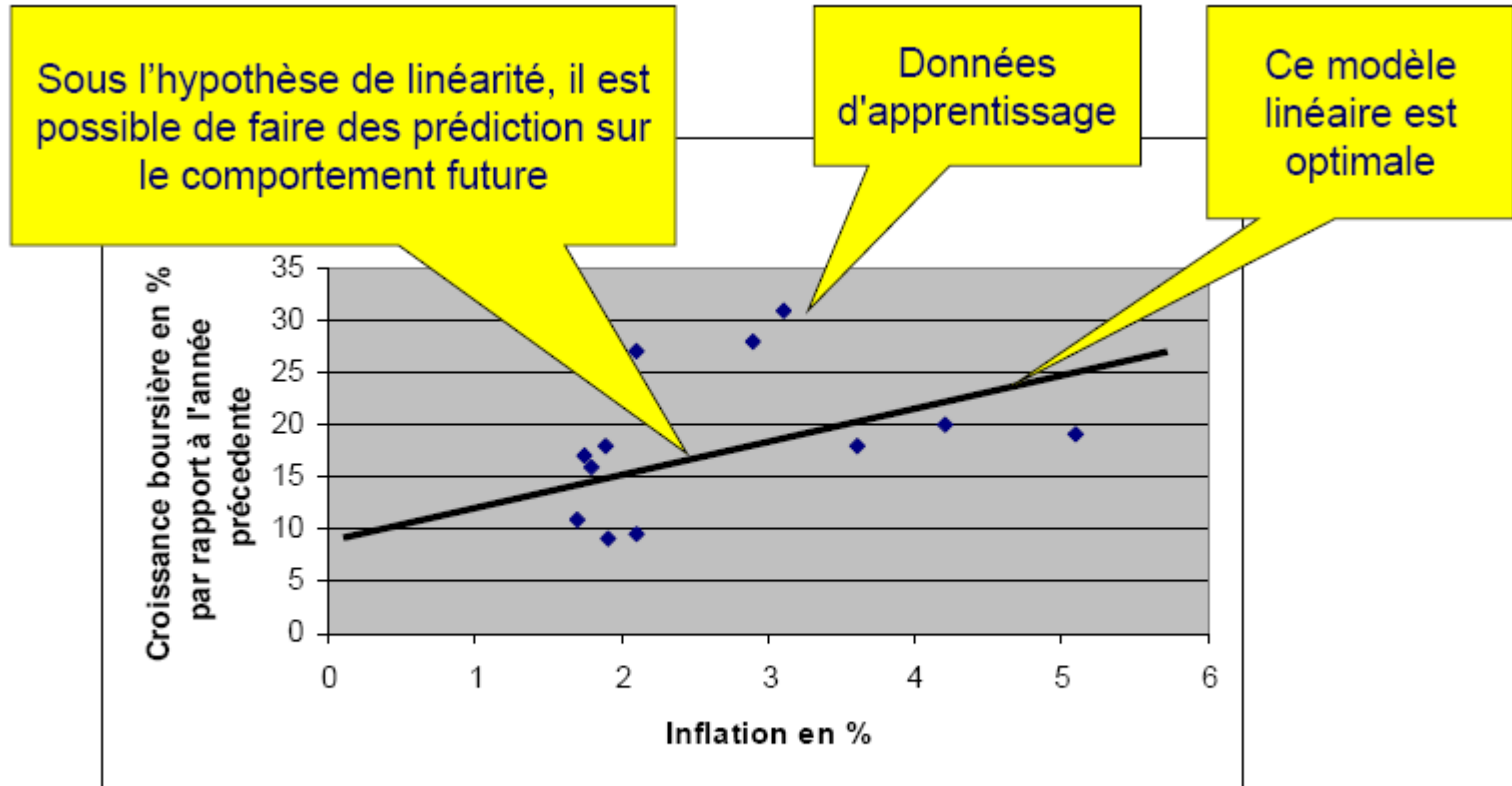


## Ensemble de test



# Exemple de modélisation (1/3)

## La régression linéaire

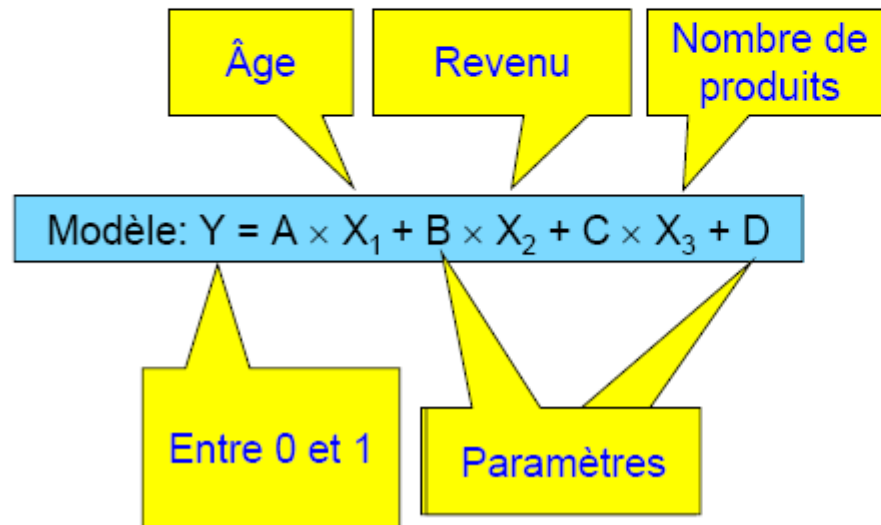


Modèle:  $Y = A \times X + B$

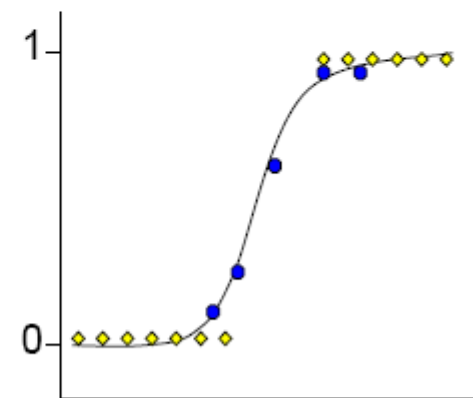
# Exemple de modélisation (2/3)

## La régression logistique

Exemple: Un ensemble comprenant deux groupes de 10'000 clients avec et sans carte de crédit. L'âge, le revenu et le nombre de produits de chaque clients sont connus (entrées). Les paramètres du modèle seront choisis (par apprentissage supervisé) tels que le modèle rende comme sortie une valeur entre 0 (client sans carte) et 1 (client avec carte).



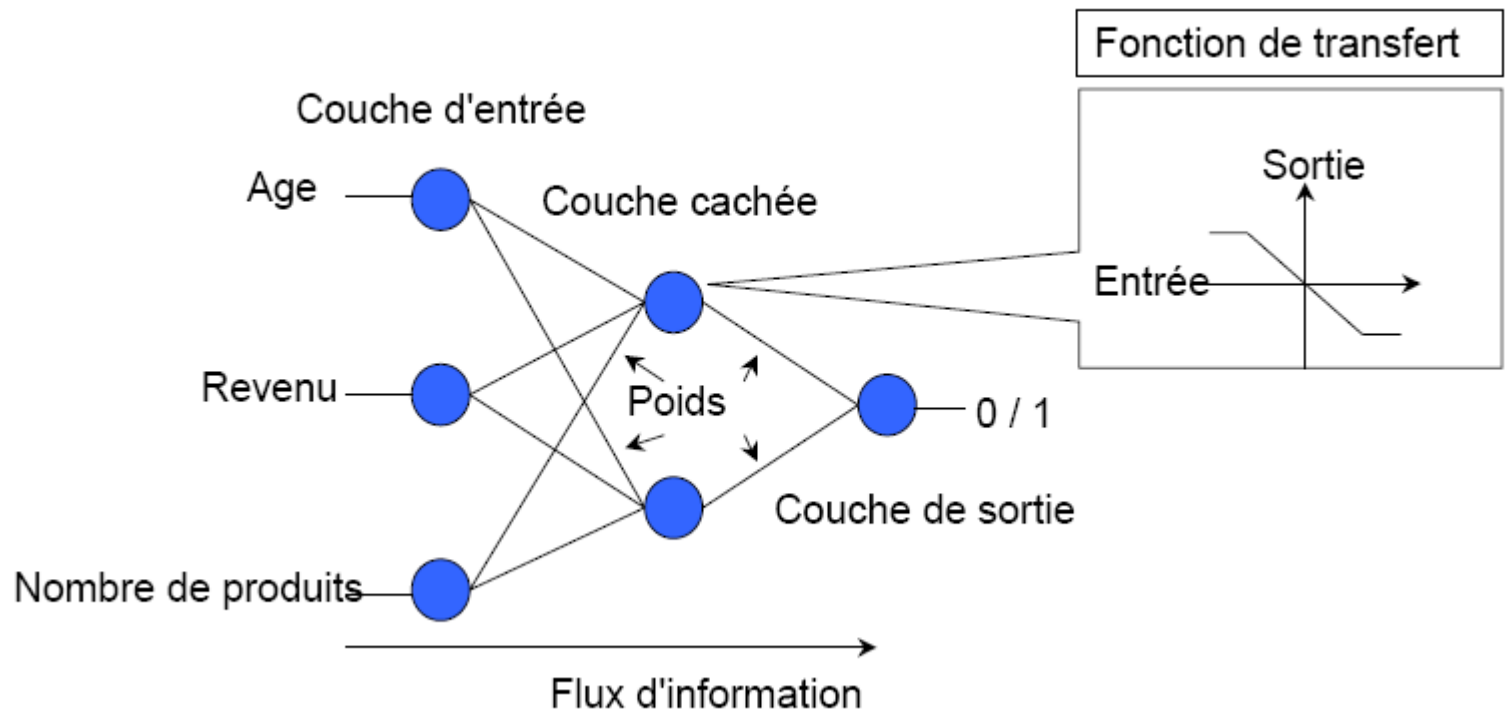
Probabilité d'achat





# Exemple de modélisation (3/3)

## Le réseau de neurones



# Les principaux objectifs stratégiques marketing

- Acquisition de nouveaux clients ou l'augmentation du rendement des actions de prospection commerciales et Marketing
- Croissance de la valeur des clients ou l'augmentation des services qui leur sont fournis
- Rétention ou fidélisation «ciblée» de la clientèle
- Win-back ou la reconquête de segments de clientèle «passés» à la concurrence
- Exploiter l'ensemble des données clientèle pour dégager des avantages concurrentiels décisifs

# Principales applications Marketing du Data Mining

## ● Les réponses du Data Mining:

- Évaluation / profilage clients
- Segmentation / typologie clients
- Score d'appétence produits / ventes croisées, détection d'opportunités de vente
- Optimisation des actions commerciales (cibles MD,...),
- Recherche d'associations, analyse de proximité, co-détention produits, «basket analysis »
- Analyse CHURN, réactivation clientèle

-...



# Avantages et résultats attendus du Data Mining

- Maximisation de l'effort commercial et adaptation aux attentes du marché
- Amélioration importante des taux de retour des actions commerciales
- Baisse significative du coût des actions Marketing
- Augmentation importante du ROI des actions
- Découverte de potentiels inexploités
- Élargissement de la clientèle «captive»
- Approche mixte business / quantitatif – Probabilité Score!

# Les principales limites du Data Mining

- Les « anciennes » méthodes restent toujours d'actualité
- Le Data Mining ne découvre pas tous les « trésors » cachés dans vos données
- Le Data Mining ne découvre pas automatiquement les informations dignes d'intérêt
- Des données de qualité décrivant les phénomènes sont indispensables
- Le Data Mining ne donne pas de réponses totalement tranchées

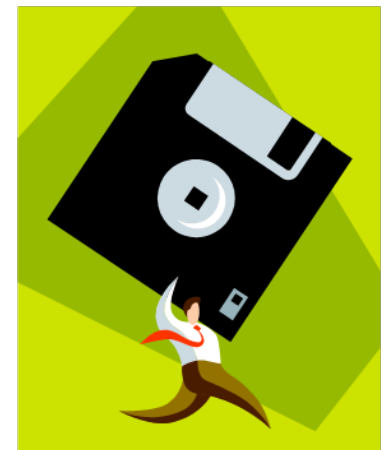
# Freins et blocages au développement du Data Mining (1/2)

## ● Au niveau « Business »

- Méconnaissance / crainte / scepticisme
- Manque de soutien du Top Management
- Engagement de la force de vente
- Difficulté à «vulgariser» certains résultats (neurones,...)
- Intégration des activités de Data Mining dans l'entreprise

## ● Au niveau des données

- Disponibilité
- Mise à jour / qualité
- Structure / historisation



# Freins et blocages au développement du Data Mining (2/2)

## ● Au niveau des outils

- Complexité des logiciels
- Spécificités des outils (intégration, fonctionnalités,...)
- Coûts liés aux applications de Data Mining

## ● Au niveau des compétences

- Absence de filière de formation «Marketing Intelligence»
- Profil complexe: Marketing, quantitatif, informatique,...



# Pistes d'améliorations pour une meilleure intégration du Data Mining

- Mise en avant des avantages concrets de la démarche
- Procéder par petites étapes successives
- Intégrer des cibles test Data Mining aux actions
- Encourager l'utilisation effective des résultats par la force de vente (effet d'entraînement)
- Travail de fond au niveau des bases de données CRM
- Les solutions logiciels en Open Source
- Développement de cours / formations spécifiques
- Sensibilisation des décideurs des entreprises



# Imposer le Data Mining par étapes

- Débuter par « récolter les fruits mûrs »
- Utilisation des données déjà disponibles
- Déploiement immédiates des résultats obtenus
- Agrandissement par incrément du système
- Intégration de la solution à l'infrastructure globale
- Intégration de sources supplémentaires
- Adaptation aux besoins du Business



# Applications concrètes du Data Mining

- Mailings commerciaux pour la vente de cartes de crédit
- Propositions de Cross Selling sur Internet
- Détection des clients prêts à quitter leur opérateur téléphonique
- Optimisation de la récolte de dons pour des œuvres humanitaires
- Détection des futurs Hits musicaux (Hit Song Science model)
- ...



# Les principaux logiciels de Data Mining

