

SOFT COMPUTING
Jamais vos clients ne vous
auront rapporté autant.

soft.computing

Data Mining : facteur de compétitivité des entreprises

Paris, le 18 Juin 2001.

René Lefebure
rlf@softcomputing.com
Soft Computing
46, rue de la Tour
75116 Paris
Tél : 01 73 00 55 00
Fax : 01 73 00 55 01
www.softcomputing.com

Sommaire

- 1. Présentation de Soft Computing**
2. *Positionnement du Data Mining*
3. *Les Apports des Techniques Nouvelles*
4. *Les apports dans l'inter-activité*
5. *Les défis du Data Mining*
6. *Références*

Soft Computing : une mission claire et précise

Structurer les données disponibles sur les clients et les marchés pour les rendre plus facilement accessibles et utilisables

- Datawarehouse

Transformer ces données en informations puis en connaissances pour guider dans les décisions et les actions

- Web mining
- Data mining
- Personnalisation

Exploiter cette connaissance pour personnaliser la relation et augmenter la valeur individuelle des clients

Valorisation du Capital Client

- e-Commerce
- Sales force automation (SFA)
- Recommandations en ligne
- Coordination multicanaux
- Gestion de campagnes (EMA)

Vue d'ensemble de la cartographie des compétences

**x-CRM :
l'entreprise
relationnelle**

**Compétences
métiers**

Stratégie

Etudes d'opportunités
Diagnostic
Analyse de la valeur
Datamining et webmining
Stratégie x-CRM

Organisation

Diagnostic des process
Définition des process
Conception de systèmes
Fiche de postes
Change Management
Stratégie marques
Opérations de recrutements
Marketing direct on&offline
Systèmes de fidélisation
Régie publicitaire

Marketing

Datawarehouse
Décisionnel
SFA et EMA
Marketplace
eCommerce
EAI et ETL
Sécurité & systèmes

Technologie

Gestion de bases
Marketing
Traitement adresses
Géomarketing
Datamining

Outsourcing

**Expertises
sectorielles**

**Services
financiers**

Retail banking
Private banking
Brokers & eBrokers
Assurances

Telecom & ISP

Opérateurs B2C
Opérateurs B2B
Providers Internet

**Distribution,
VPC**

Hyper et super
Distribution spécialisée
VPC B2C
VPC B2B
eCommerce

Industries

Automobile
Chimie
Utilities

**Tourisme
& transport**

Tour opérateurs
Parcs de loisirs
Transporteurs
Compagnies aérienne

Santé & PGC

Pharma & cosméto
Electronique
Agro-alimentaire
Edition musique &
Presse

**Maîtrise des
technologies**

Face to face

Siebel
Vantive
Onyx

eCommerce

Commerce Serv
Websphere
Broadvision
ePiphany

Call center

Conso+
Point

Marketplace

Webmethods
Ariba

**Database
marketing**

Datastage
Oracle Teradata
Sqlserver
Prime, marketic
Sas, IBM, Spss

Une expertise formalisée

- Un positionnement en expertise marketing



Le Data Mining
Eyrolles 2001
Prix du meilleur livre
informatique,
décerné par l'AFISI

Gestion de la
Relation Client
Eyrolles 2000



Sommaire

1. *Présentation de Soft Computing*
2. **2. Positionnement du Data Mining**
3. *Les Apports des Techniques Nouvelles*
4. *Les apports dans l'inter-activité*
5. *Les défis du Data Mining*
6. *Références*

L'opposition exploratoire et confirmatoire

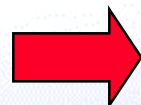
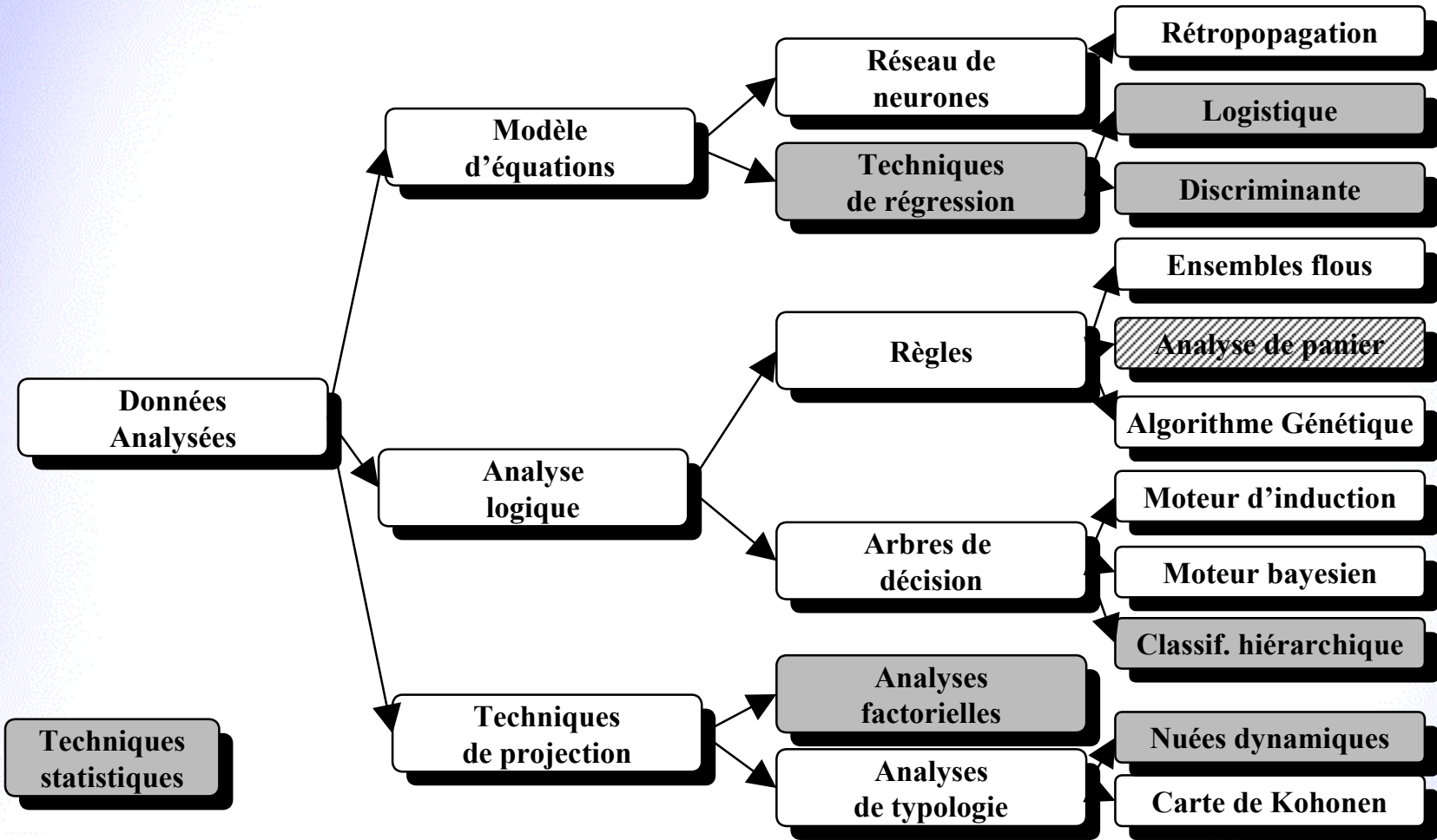
● Approche confirmatoire

- Poser une hypothèse : H_1/H_0
- Valider l'hypothèse avec les tests statistiques
- Techniques statistiques
- **Problème :**
 - avoir du temps....
 - et de l'imagination !

● Approche exploratoire

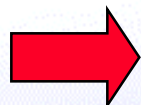
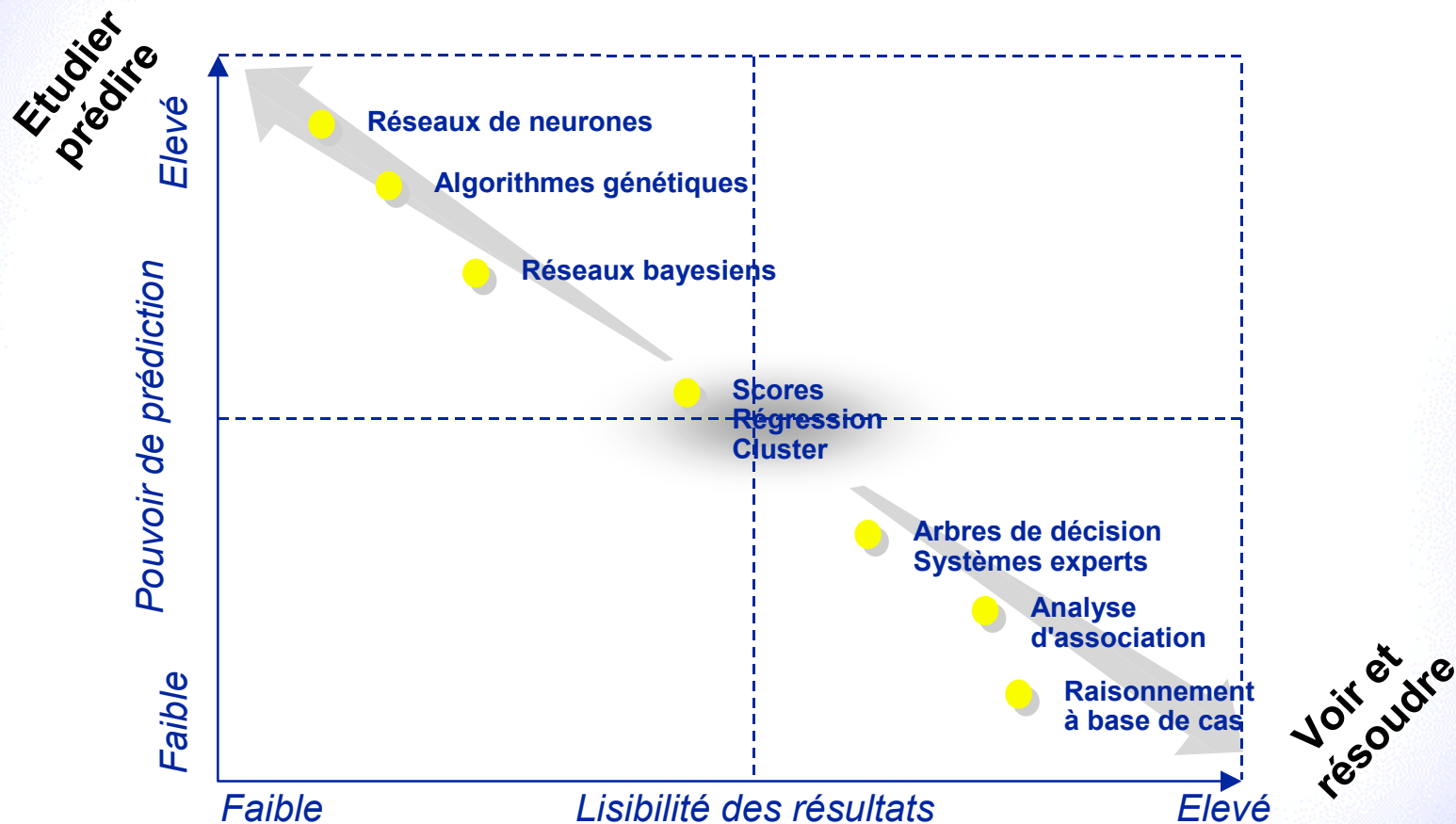
- Collecter des données
- Explorer l'espace des dimensions
- Techniques inductives
- **Problème :**
 - savoir naviguer dans les données... et les techniques !

Le positionnement entre les statistiques et les outils d'analyse automatique



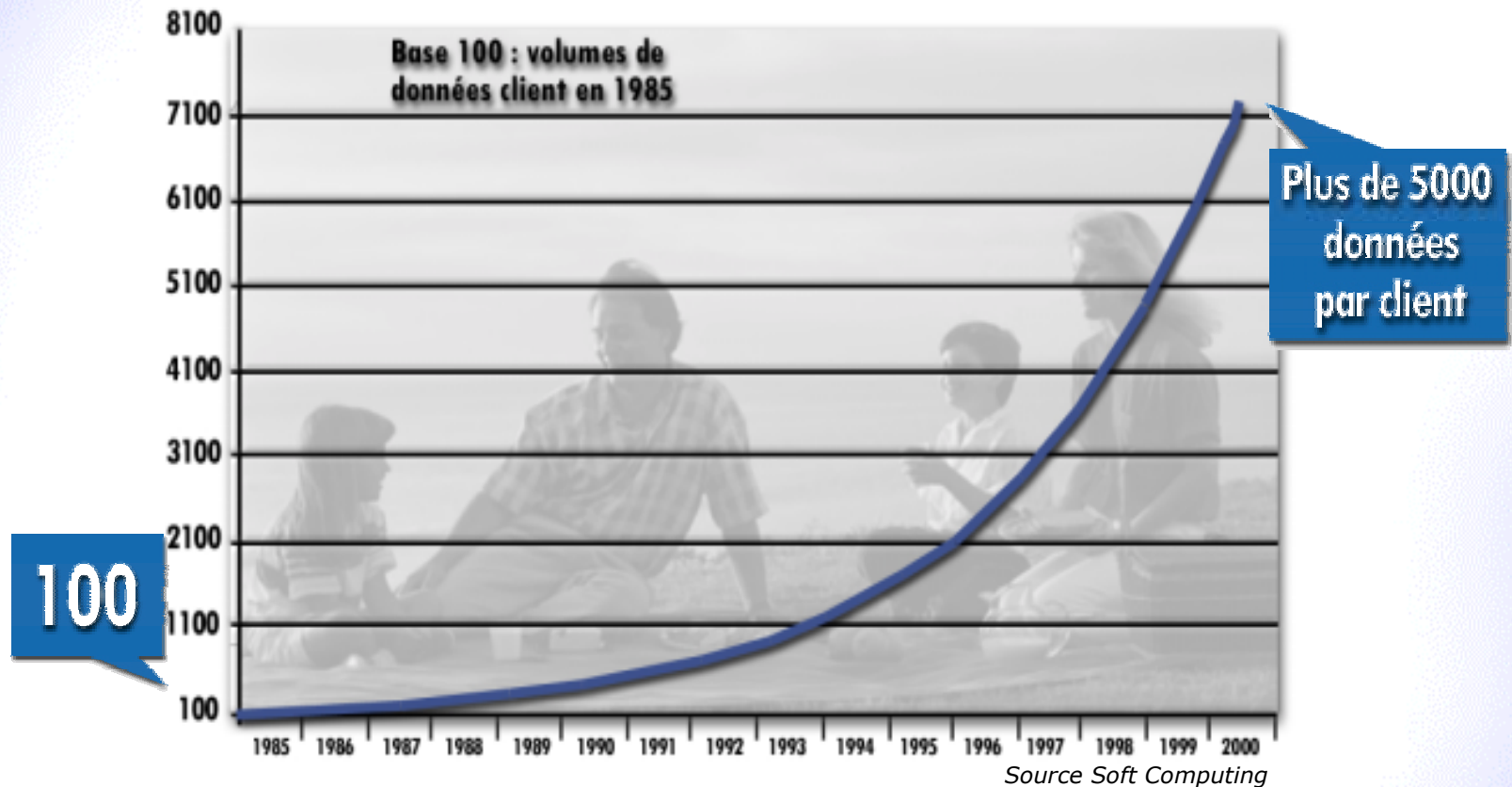
Il est toujours possible de répondre à une même problématique par plusieurs techniques !

Le choix des techniques ne dépend pas que des critères de performance



Il est souvent plus efficace d'avoir un modèle sous-optimal dans certains cas !

Un constat : une explosion des données client



La multiplicité des moyens de collecte de l'information (Internet, call center, contact direct, mailing, enquête ...) et les capacités de stockage favorisent la croissance du volume de données qui se rapporte à chaque client.

Qui concerne l'ensemble des secteurs d'activités



Le volume des données est doublé tous les 18 à 24 mois.

Échapper à l'utopie

Données disponibles :
+150 % / an

Productivité actuelle : 2j / ciblage
8j / analyse factorielle...



En 2015, les statisticiens représenteraient plus de 50% de la population active !!

Besoins d'études :
+45 % / an

Les algorithmes de data mining sont une réponse au besoin croissant d'industrialisation des études de données

Spécificités du data mining

- Non limitées aux données numériques : les nouvelles techniques permettent de traiter des types de données différents : son, texte, image, objet.
- Capacité d'auto-organisation : les nouveaux algorithmes sont des processus auto-organiseurs guidés par les données, alors que les statistiques s'appuient sur un processus de vérification d'hypothèses guidé par un utilisateur.
- Facilité d'utilisation : une partie des nouveaux algorithmes peut être utilisée par des utilisateurs métiers dans la résolution de problèmes et l'aide à la décision.
- Puissance de recherche : les nouveaux algorithmes parcourent de manière automatique l'ensemble des combinaisons possibles.
- Scalabilité : les nouveaux algorithmes sont optimisés pour traiter des volumes importants de données (enregistrements et variables) en phase d'analyse.

Sommaire

1. *Présentation de Soft Computing*
2. *Positionnement du Data Mining*
- 3. *Les Apports des Techniques Nouvelles***
4. *Les apports dans l'inter-activité*
5. *Les défis du Data Mining*
6. *Références*

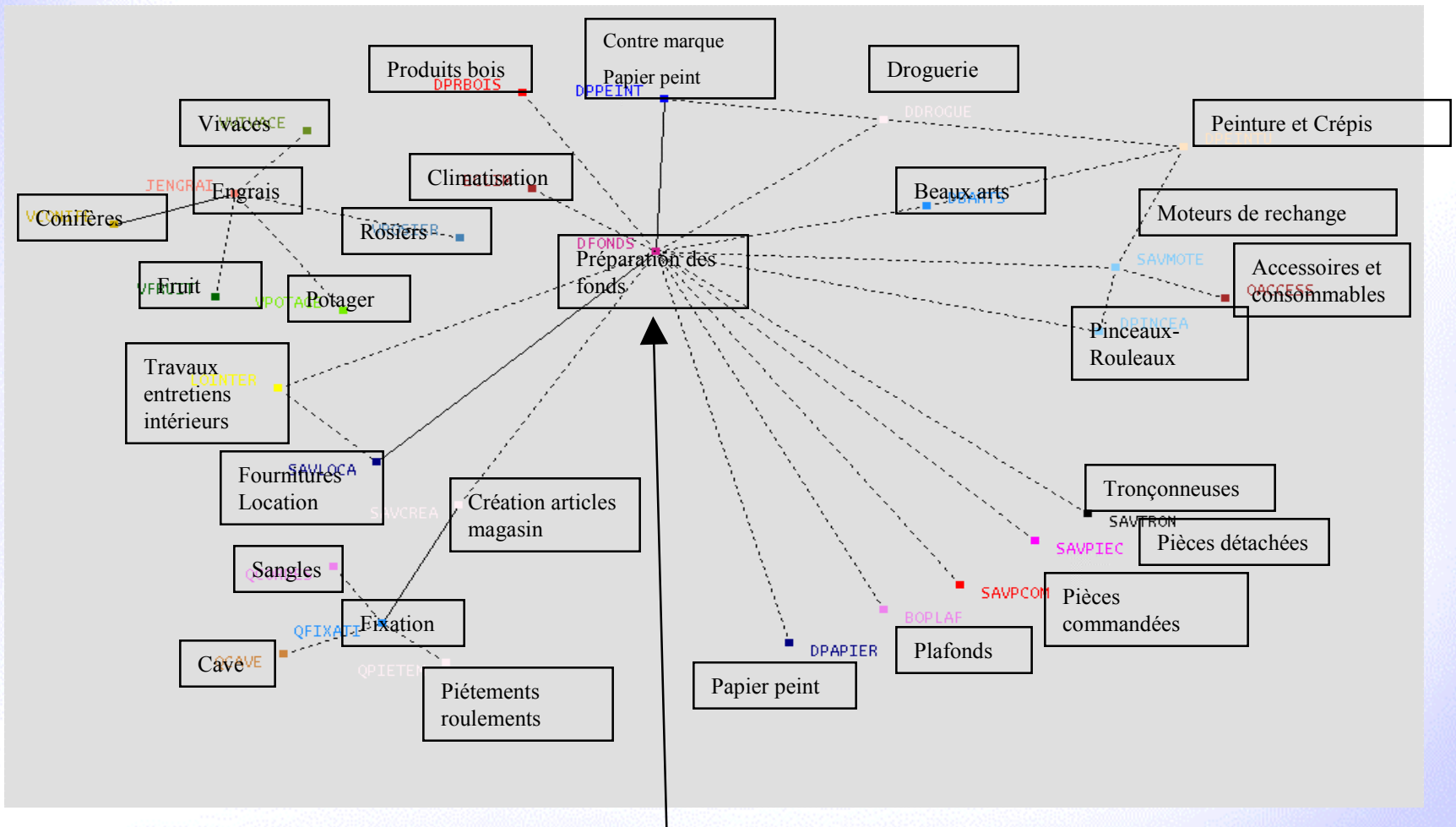
Les analyses d'association

Identification des opportunités de ventes croisées

1	Famille A	Famille B	EFFECTIF	CONFIANCE de la FAMILLE A	LIFT	SUPPORT
2	Bati-Center	Bati-Center	5960	100.00	1.49	67.28
3	Bati-Center	Service après vente	1429	23.98	1.31	16.13
4	Bati-Center	Location de matériels	299	5.02	1.30	3.38
5	Bati-Center	Végétaux	3382	56.74	1.16	38.18
6	Bati-Center	Bois-Panneaux	4637	77.80	1.14	52.35
7	Bati-Center	Décoration	5252	88.12	1.06	59.29
8	Bati-Center	Électricité	5019	84.21	1.09	56.66
9	Bati-Center	JARDIN	4591	77.03	1.12	51.83
10	Bati-Center	Outillage	4881	81.90	1.11	55.10
11	Bati-Center	Quincaillerie	5385	90.35	1.07	60.79
12	Bati-Center	Revetements de sols	3495	58.64	1.18	39.46
13	Bati-Center	Sanitaire	4720	79.19	1.12	53.29
14	Bois-Panneaux	Bati-Center	4637	76.91	1.14	52.35
15	Bois-Panneaux	Service après vente	1351	22.41	1.22	15.25
16	Bois-Panneaux	Location de matériels	281	4.66	1.21	3.17
17	Bois-Panneaux	Végétaux	3352	55.60	1.13	37.84
18	Bois-Panneaux	Bois-Panneaux	6029	100.00	1.47	68.06
19	Bois-Panneaux	Décoration	5375	89.15	1.08	60.68
20	Bois-Panneaux	Électricité	5032	83.46	1.08	56.81
21	Bois-Panneaux	JARDIN	4451	73.83	1.08	50.25
22	Bois-Panneaux	Outillage	4837	80.23	1.09	54.61
23	Bois-Panneaux	Quincaillerie	5512	91.42	1.08	62.23
24	Bois-Panneaux	Revetements de sols	3545	58.80	1.19	40.02
25	Bois-Panneaux	Sanitaire	4706	78.06	1.10	53.13
26	Décoration	Bati-Center	5252	71.61	1.06	59.29
27	Décoration	Service après vente	1525	20.79	1.13	17.22
28	Décoration	Location de matériels	314	4.28	1.11	3.54
29	Décoration	Végétaux	3917	53.41	1.09	44.22
30	Décoration	Bois-Panneaux	5375	73.29	1.08	60.68
31	Décoration	Décoration	7334	100.00	1.21	82.80

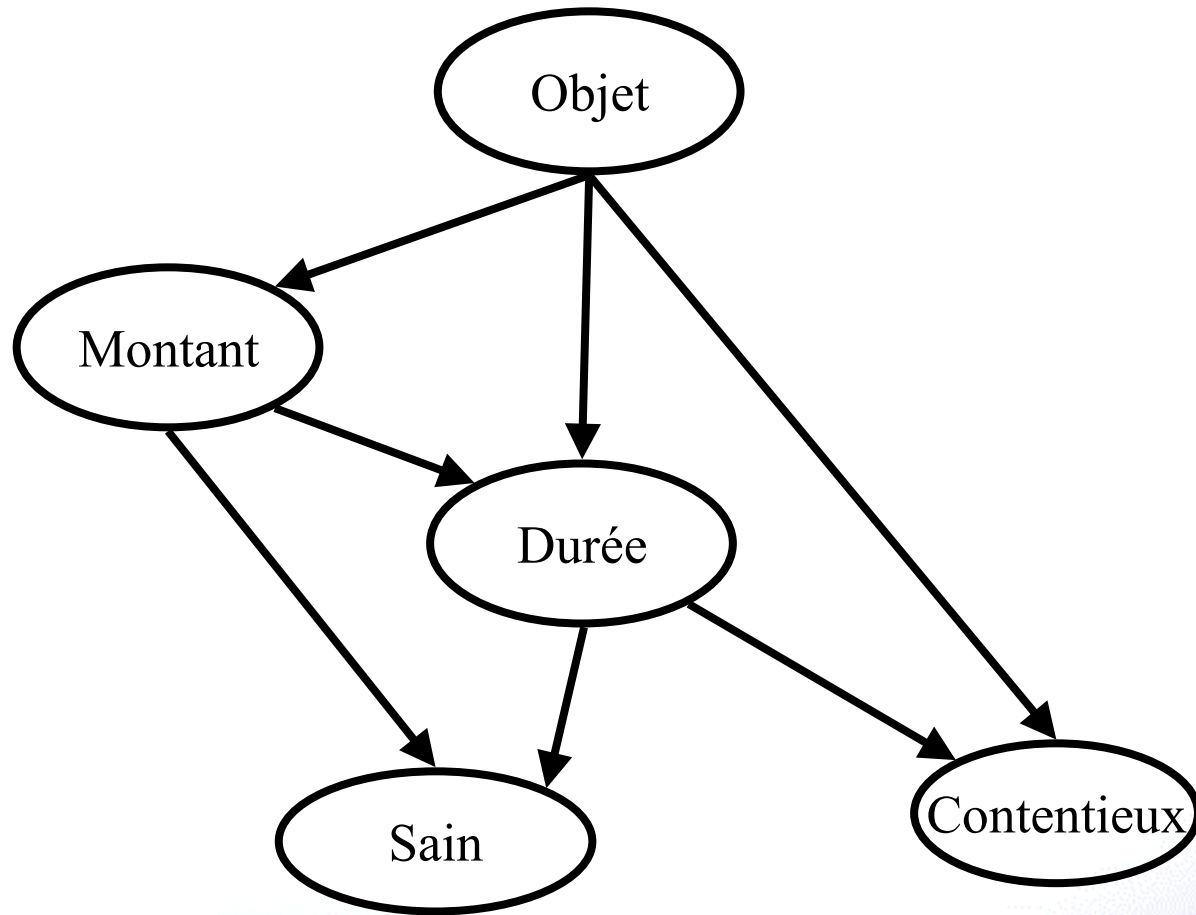
comptage / association / tri / liste / modeles

Les « toiles » d'association



Typique d'un rayon rayonnant

Les réseaux bayésiens



Les recherches de causalité en étude de satisfaction

Schéma sans relation causale

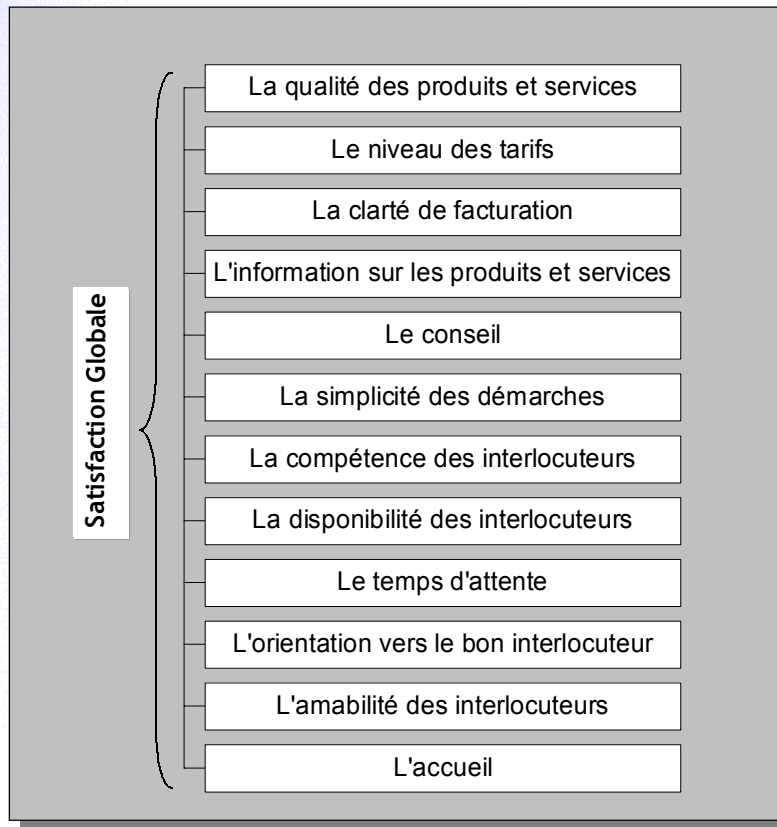
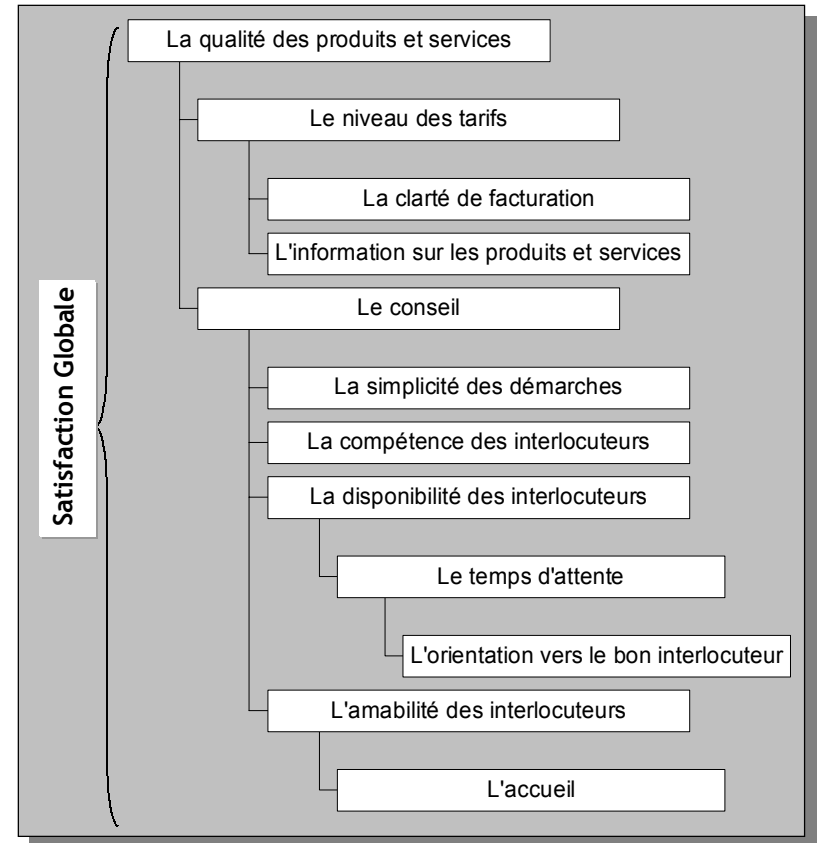
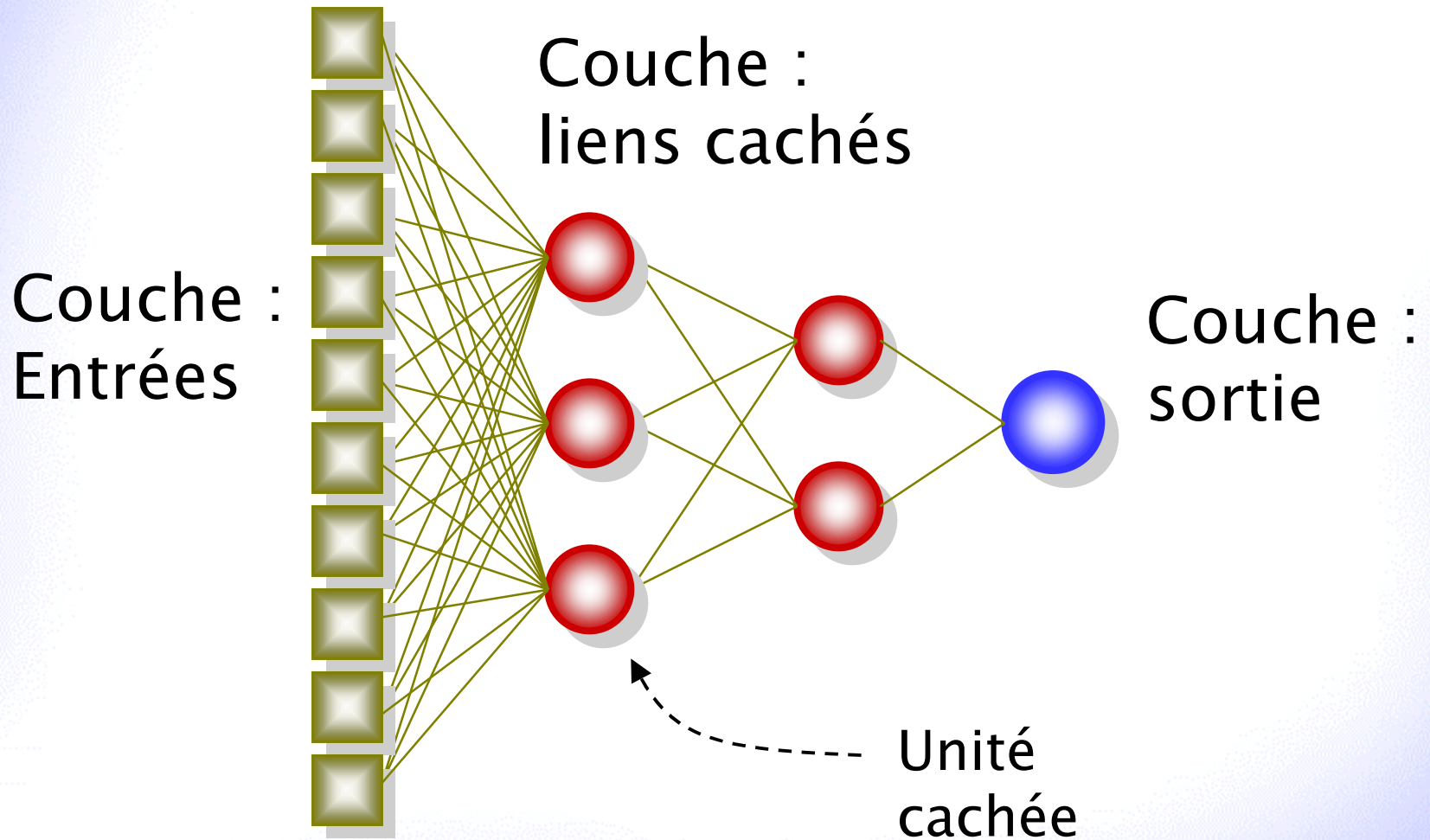


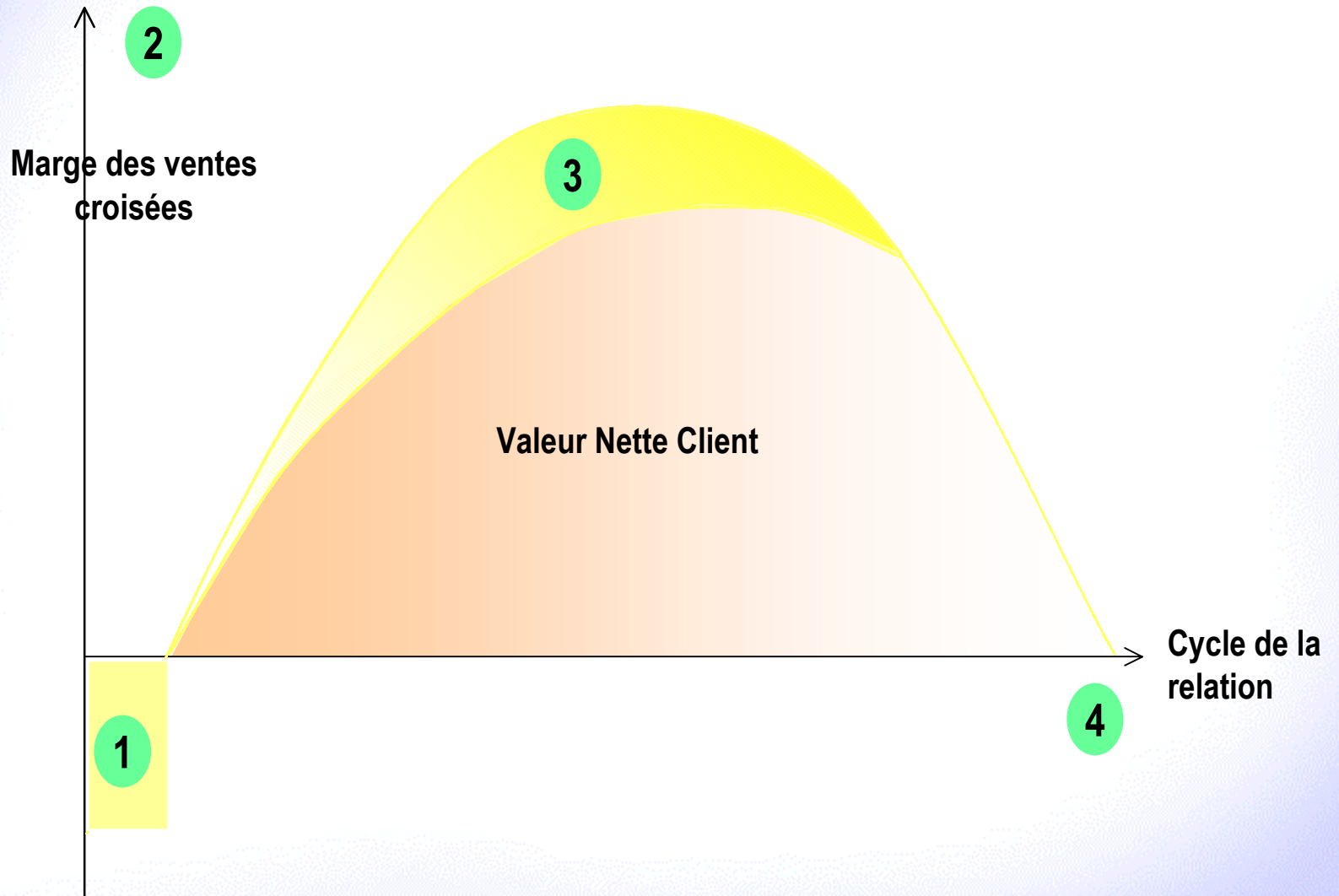
Schéma causal



Les réseaux de neurones

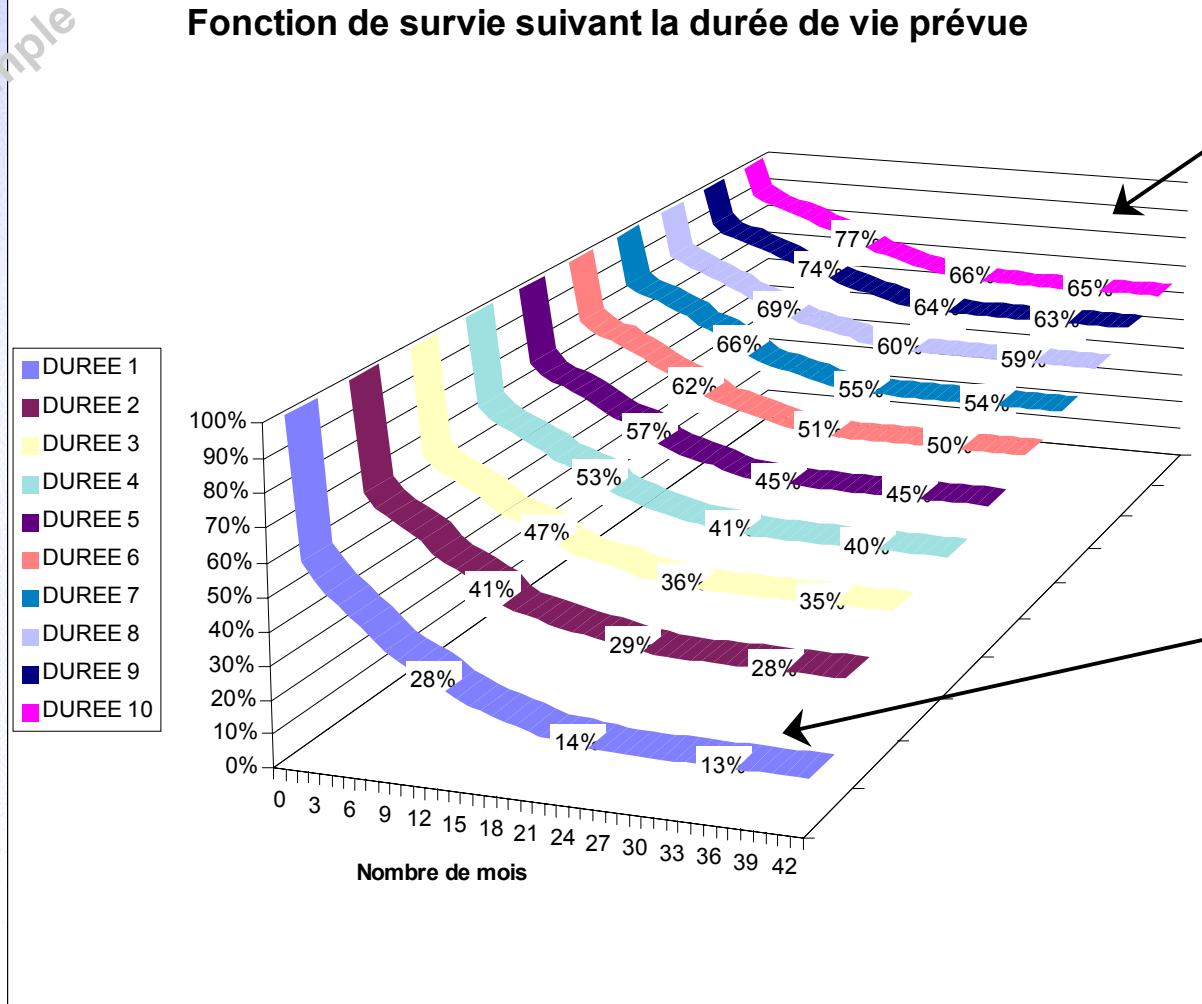


La détermination de la Life Time Value



La validation de la durée sur l'historique

Fonction de survie suivant la durée de vie prévue

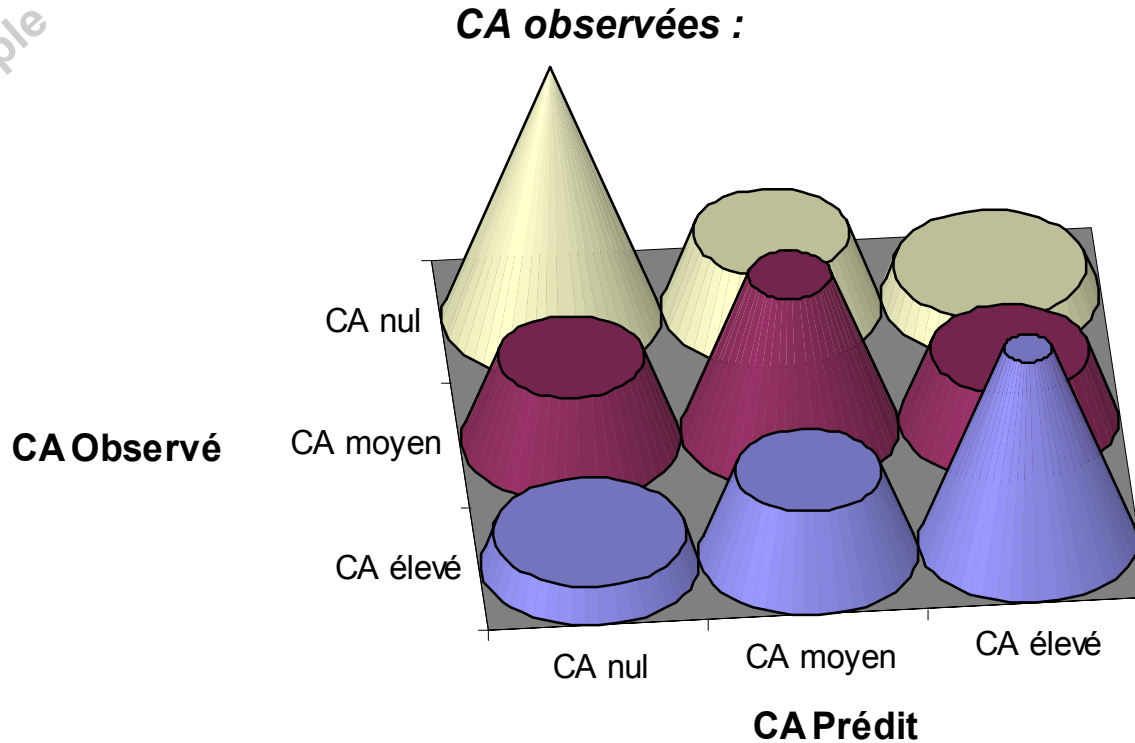


65% des cartes créées en 1997 pour lesquelles nous avons prévu une durée de vie très longue, sont encore en activité au début de leur 3ème année de vie

13% des cartes créées en 1997 pour lesquelles nous avons prévu une durée de vie très courte, sont encore en activité au début de leur 3ème année de vie
(72% sont mortes après 12 mois d'activité, 86% après 24 mois)

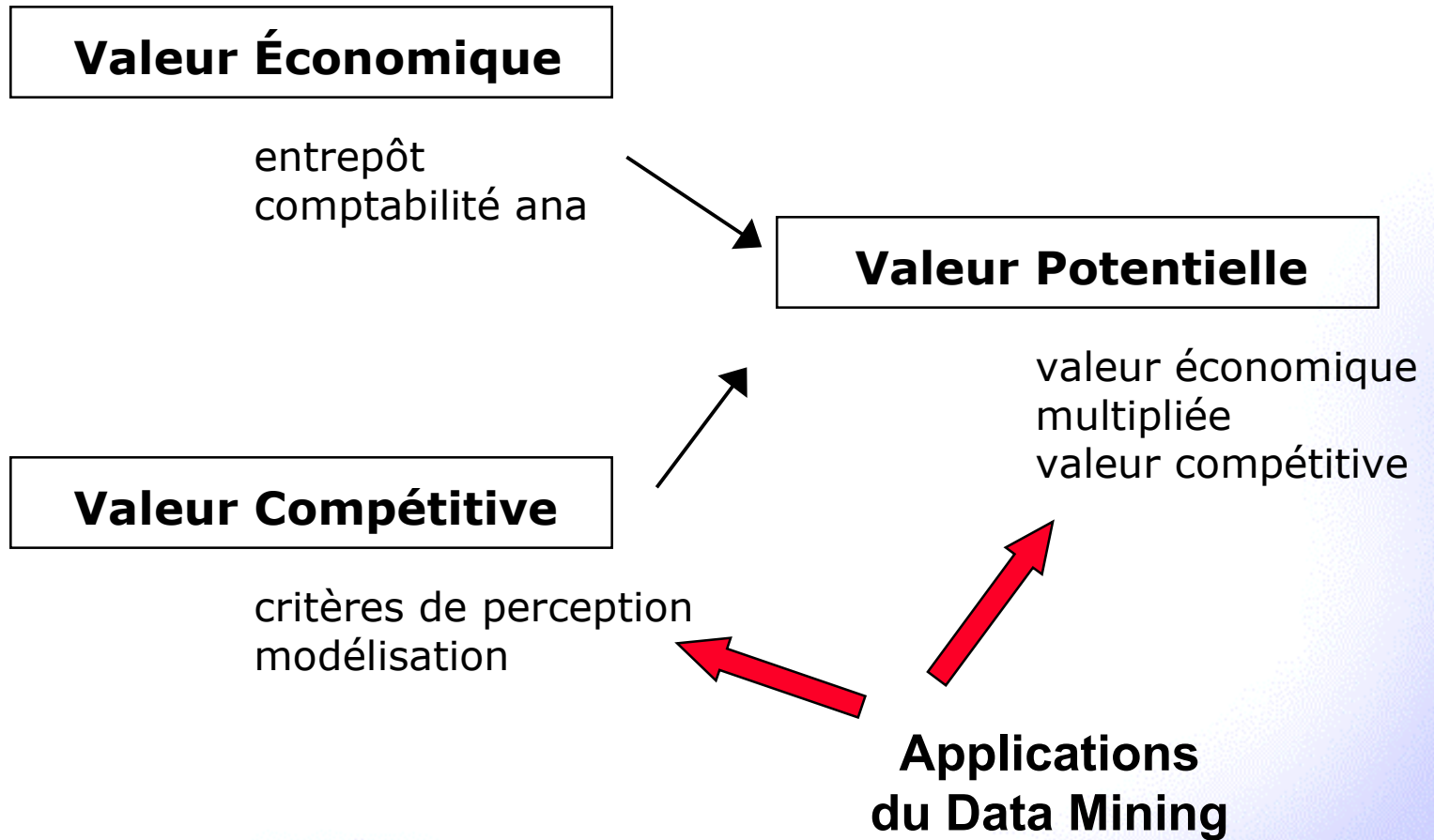
La détermination du potentiel

Reclassement des porteurs par rapport aux tranches de CA observées :

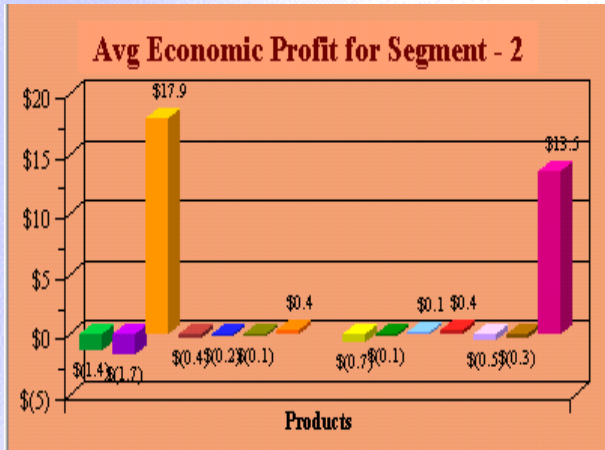


On distingue bien la diagonale, signe que le reclassement des clients dans leur tranche d'origine est fiable.

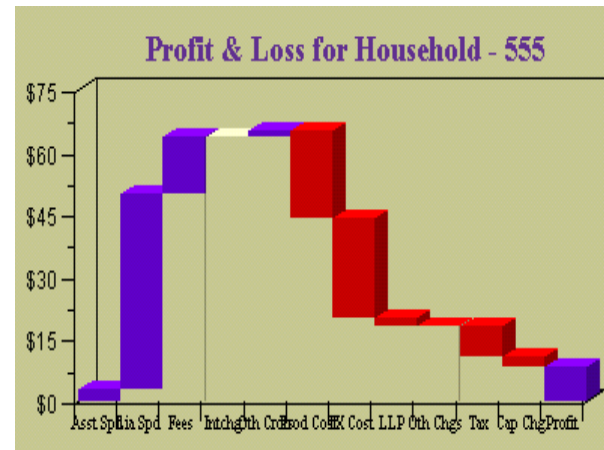
A la poursuite de la valeur client



L'optimisation de la valeur client



Comment augmenter les revenus ?



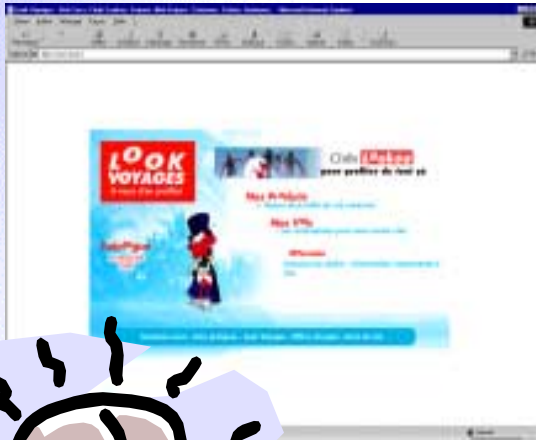
Comment diminuer les coûts ?

La multiplication des canaux et des moyens d'interactions avec le client posera de manière de plus en plus forte la pertinence du contenu et la justification économique des investissements marketing

Sommaire

1. *Présentation de Soft Computing*
2. *Positionnement du Data Mining*
3. *Les Apports des Techniques Nouvelles*
- 4. *Les apports dans l'inter-activité***
5. *Les défis du Data Mining*
6. *Références*

La constitution des fichiers logs



Chaque action sur le site correspond à une requête HTTP (click, bandeau de publicité,...) qui est envoyée au serveur Web



Serveur WEB Look

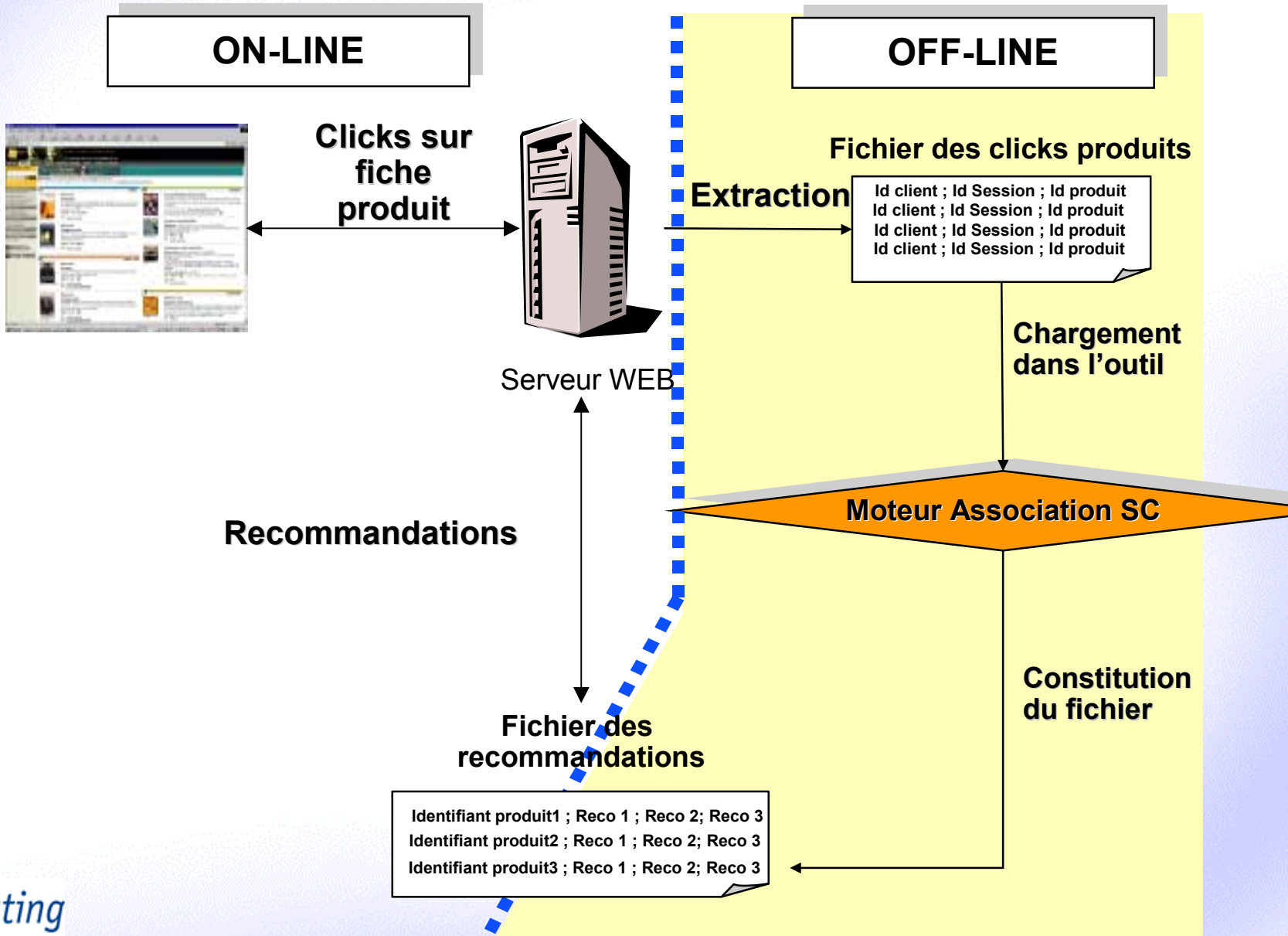
Ces actions sont stockées sur un fichier texte dont l'extension est .log d'où le nom de fichiers log

```
2000-05-19 20:28:31 195.127.1.190 - W3SVC1 JCH 195.127.1.199 80 GET /images/logoSC.gif - 304 0 140 439 0 HTTP/1.1 jch
Mozilla/4.0+(compatible;+MSIE+5.01;+Windows+NT+5.0)
time=10%3A58%3A32;+date=18%2F05%2F2000;+vote=195%2E127%2E1%2E199;+Nom=jc;+ASPSESSIONIDQQGGRDC=PAAHPBNCAP
GAEBGPCDGLHJKK http://jch/

2000-05-19 21:15:27 195.127.1.190 - W3SVC1 JCH 195.127.1.199 80 POST /chercher.asp - 200 0 89 597 151 HTTP/1.1 jch
Mozilla/4.0+(compatible;+MSIE+5.01;+Windows+NT+5.0)
time=10%3A58%3A32;+date=18%2F05%2F2000;+vote=195%2E127%2E1%2E199;+Nom=jc;+ASPSESSIONIDQQGGRDC=BBAHPBNCNH
EDPGIIGLLDGGIA http://jch/
```

```
#Fields: date time c-ip cs-username s-sitename s-computername s-ip s-port cs-method cs-uri-stem cs-uri-query sc-status sc-win32-status sc-bytes cs-bytes time-taken cs-version cs-host cs(User-Agent) cs(Cookie) cs(Referer)
```

L'architecture du système de recommandations

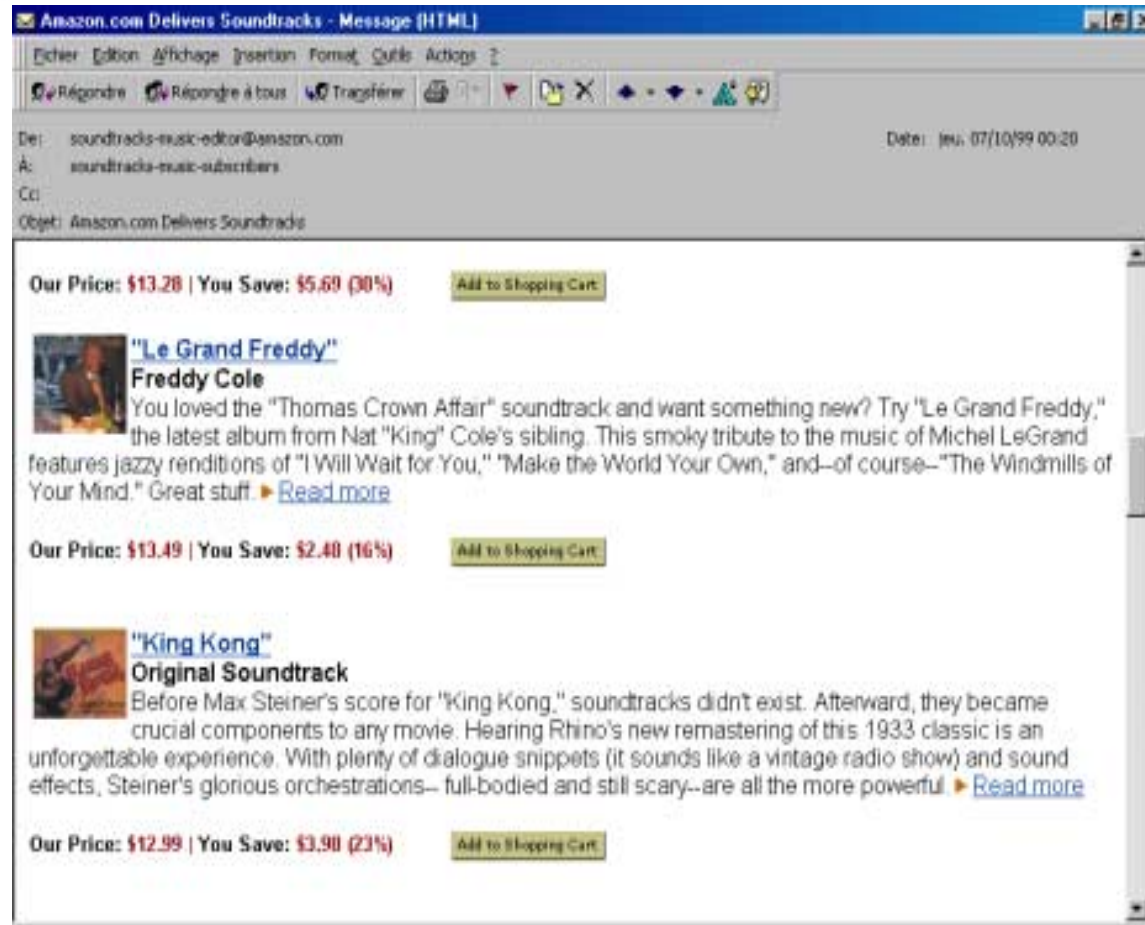


Les associations sur Amazon.com

- La capacité d'associer dynamiquement des produits complémentaires : cross selling
- La capacité de proposer des produits adaptés aux centres d'intérêts du client dès la première page.

The screenshot displays the Amazon.com Similarity Explorer interface. The browser window title is "Amazon.com Similarity Explorer - Microsoft Internet Explorer". The address bar shows the URL: http://www.amazon.com/exec/obidos/tg/sim-explorer/explorer-items/-/1061002777/D/ref=pd_sim_d_b/104-7982137-8711936. The page features the Amazon.com logo and navigation links like "WELCOME", "DIRECTORY", "YOUR ACCOUNT", and "HELP". A search bar is present with "All Products" selected. The main content area is titled "Recommendations > Similarity Explorer" and shows "People who bought this item...". The featured product is "Professional Java Server Programming: with Servlets, JavaServer Pages (JSP), XML, Enterprise JavaBeans (EJB), JNDI, CORBA, Jini and Javaspaces" by Andrew Patzer, et al. The price is \$47.99. Below this, there are "Add to Cart" and "Add to Wish List" buttons. The "Also bought these items..." section shows three related products: "Java Servlet Programming" by Jason Hunter, et al; "Enterprise JavaBeans" by Richard Monson-Haefel; and "Developing Java Servlets" by James Goodwill. A "Refresh" button is located on the left side of this section. The footer of the browser window shows the URL: http://www.amazon.com/exec/obidos/subst/home/home.html/ref=hp_gateway/104-7982137-8711936.

Les campagnes Email sur des associations




Amazon.com Delivers Soundtracks - Message (HTML)

Fichier Edition Affichage Insertion Format Outils Actions


Répondre Répondre à tous Transférer

De: soundtracks-music-editor@amazon.com Date: Jeu, 07/10/99 00:20
À: soundtracks-music-subscribers
Cc:
Objet: Amazon.com Delivers Soundtracks

Our Price: \$13.28 | You Save: \$5.69 (30%) [Add to Shopping Cart](#)

 **"Le Grand Freddy"**
Freddy Cole
You loved the "Thomas Crown Affair" soundtrack and want something new? Try "Le Grand Freddy," the latest album from Nat "King" Cole's sibling. This smoky tribute to the music of Michel LeGrand features jazzy renditions of "I Will Wait for You," "Make the World Your Own," and--of course--"The Windmills of Your Mind." Great stuff. ▶ [Read more](#)

Our Price: \$13.49 | You Save: \$2.40 (16%) [Add to Shopping Cart](#)

 **"King Kong"**
Original Soundtrack
Before Max Steiner's score for "King Kong," soundtracks didn't exist. Afterward, they became crucial components to any movie. Hearing Rhino's new remastering of this 1933 classic is an unforgettable experience. With plenty of dialogue snippets (it sounds like a vintage radio show) and sound effects, Steiner's glorious orchestrations-- full-bodied and still scary--are all the more powerful. ▶ [Read more](#)

Our Price: \$12.99 | You Save: \$3.90 (23%) [Add to Shopping Cart](#)

Sommaire

1. *Présentation de Soft Computing*
2. *Positionnement du Data Mining*
3. *Les Apports des Techniques Nouvelles*
4. *Les apports dans l'inter-activité*
- 5. *Les défis du Data Mining***
6. *Références*

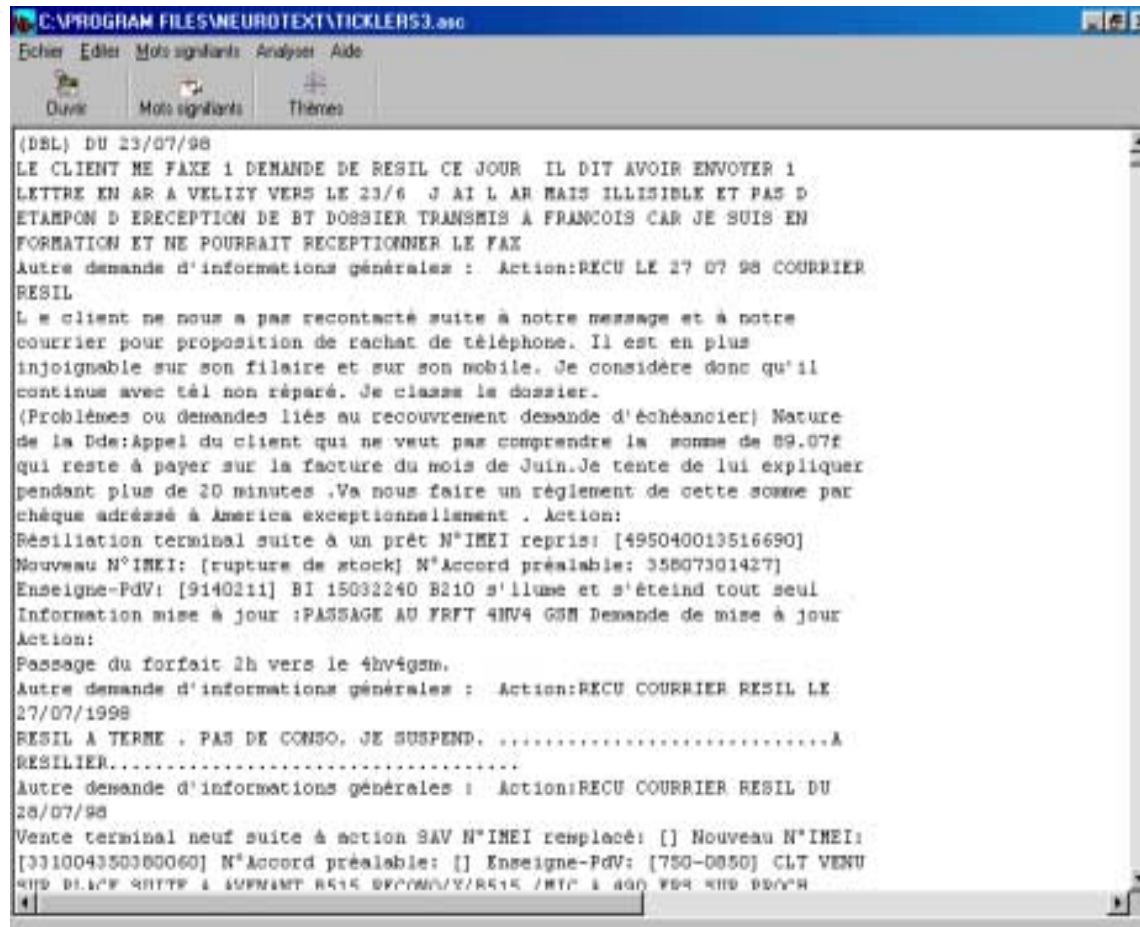
La réduction de la phase de préparation des données

The screenshot displays the AMADEA - LogWeb software interface, titled "[4.Agrégats par semaine]". The main workspace shows a complex data processing pipeline with the following components:

- Inputs:** "Logs session" and "Nombre de semaines".
- Transformations:**
 - "Regroupement par semaine" (Grouping by week)
 - "Aggregation" (Aggregation)
 - "Moyenne du nombre de sessions" (Average number of sessions)
 - "Comptage pour la semaine du reporting" (Counting for the reporting week)
 - "Moyenne pour le reporting" (Average for reporting)
 - "Trafic par semaine et par format de fichiers" (Traffic by week and file format)
 - "Trafic pour le semaine de reporting" (Traffic for the reporting week)
 - "Trafic moyen" (Average traffic)
- Operations:** Multiple "Join" operations connecting the transformed data streams.
- Output:** "Create File" (Créer un fichier).

The interface includes a "Standard" toolbar with icons for Union, Join, Selection, Create Field, Aggregation, and Create link. A "Connections" tree on the left shows the project structure, with "4.Agrégats par semaine" selected. The right sidebar contains an "Info" panel and a description of the aggregation process. The Windows taskbar at the bottom shows the system tray with the time 09:05 and the AMADEA - LogWeb application icon.

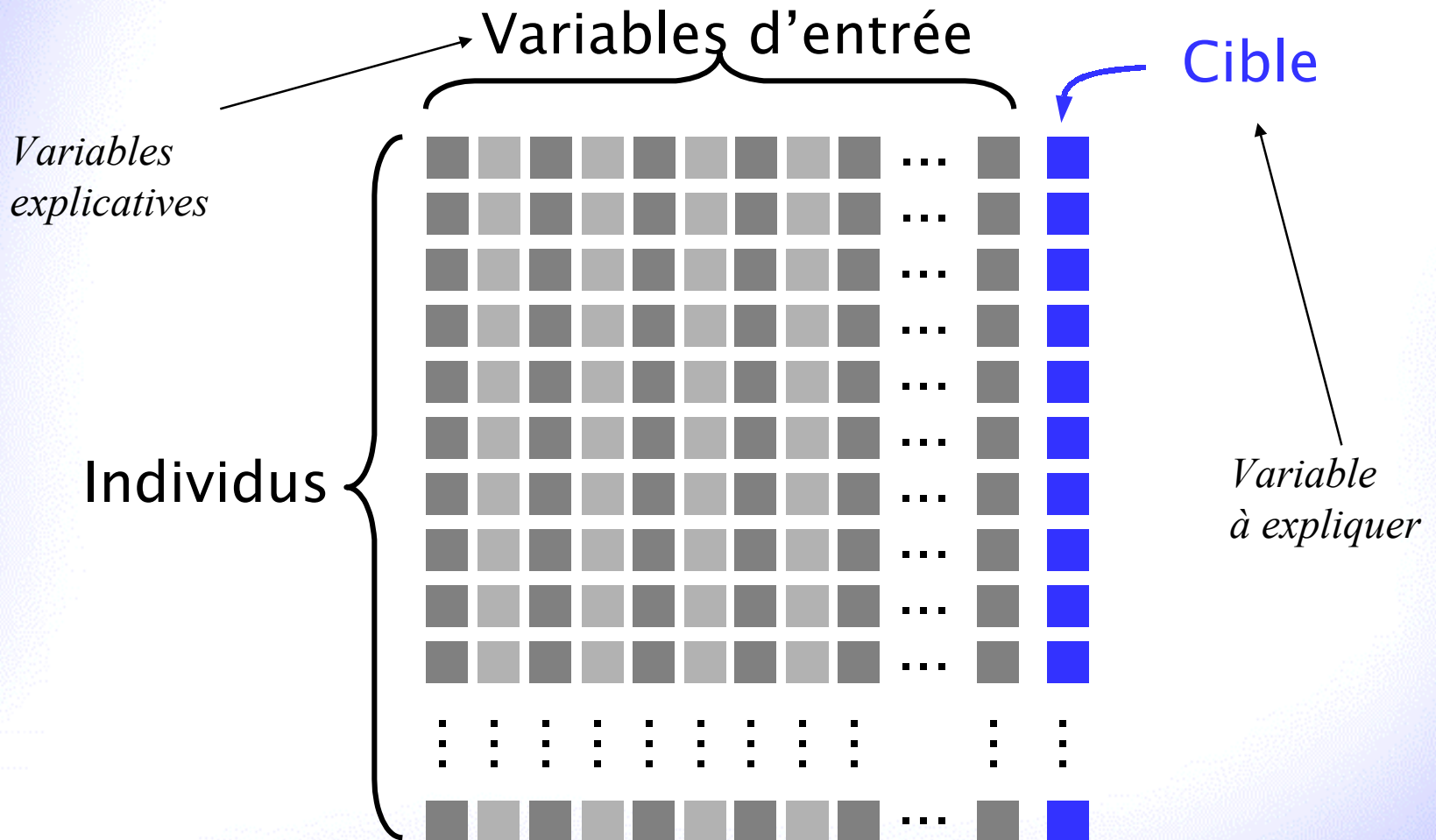
L'intégration des données textuelles



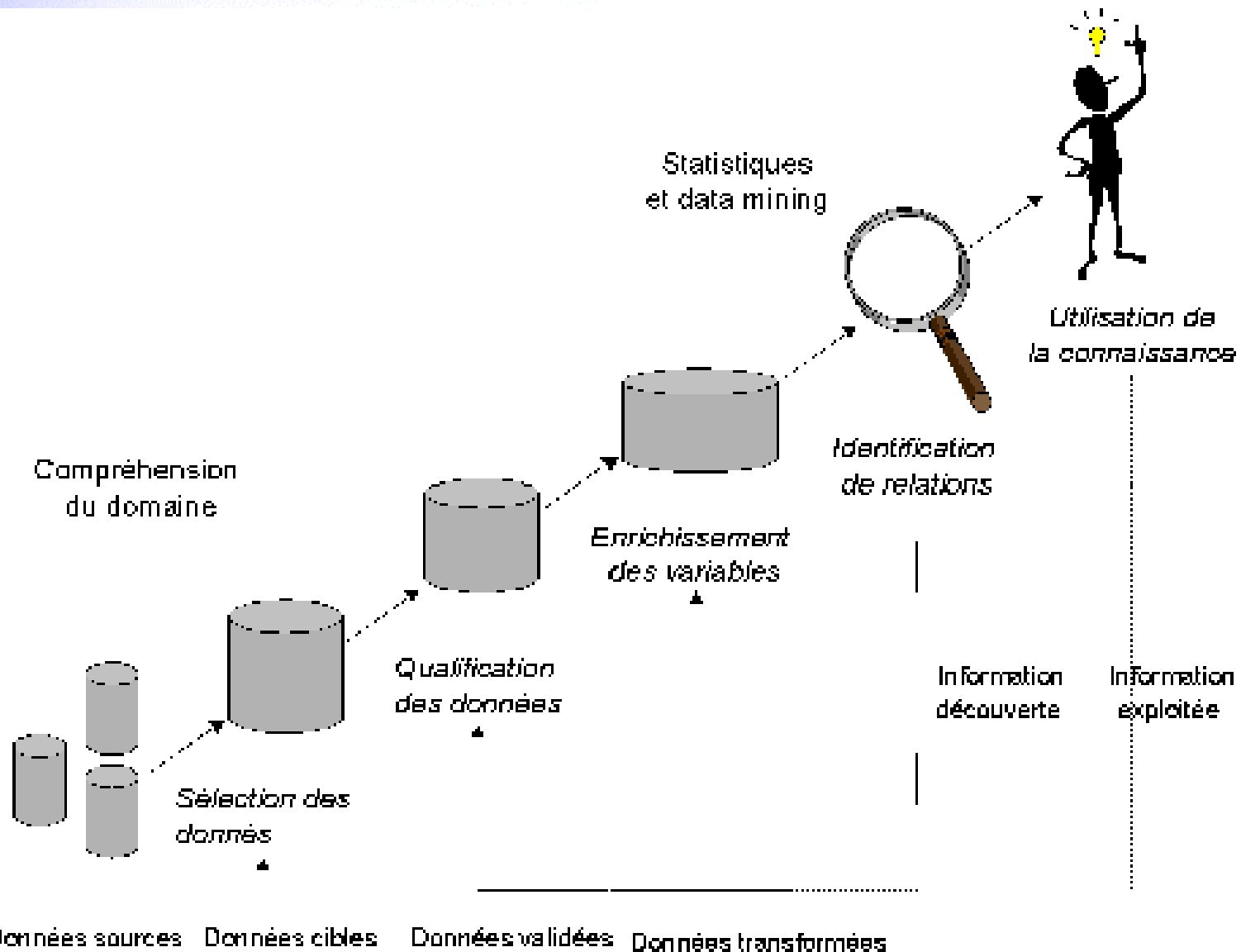
```
C:\PROGRAM FILES\NEUROTEXTATICKLERS3.aso
Fichier Editer Mots signifiants Analyse Aide
Ouvrir Mots signifiants Thèmes

(DBL) DU 23/07/98
LE CLIENT ME FAXE 1 DEMANDE DE RESIL CE JOUR IL DIT AVOIR ENVOYER 1
LETTRE EN AR A VELIZY VERS LE 23/6 J AI L AR MAIS ILLISIBLE ET PAS D
ETAMPON D ERECEPTION DE BT DOSSIER TRANSMIS A FRANCOIS CAR JE SUIS EN
FORMATION ET NE POURRAIT RECEPIONNER LE FAX
Autre demande d'informations générales : Action:RECU LE 27 07 98 COURRIER
RESIL
Le client ne nous a pas recontacté suite à notre message et à notre
courrier pour proposition de rachat de téléphone. Il est en plus
injoignable sur son filaire et sur son mobile. Je considère donc qu'il
continue avec tél non réparé. Je classe le dossier.
(Problèmes ou demandes liés au recouvrement demande d'échéancier) Nature
de la Dde:Appel du client qui ne veut pas comprendre la somme de 89.07f
qui reste à payer sur la facture du mois de Juin.Je tente de lui expliquer
pendant plus de 20 minutes .Va nous faire un règlement de cette somme par
chèque adressé à America exceptionnellement . Action:
Résiliation terminal suite à un prêt N°INEI repris: [495040013516690]
Nouveau N°INEI: [rupture de stock] N°Accord préalable: 35807301427]
Enseigne-PdV: [9140211] BI 15032240 B210 s'illum et s'éteind tout seul
Information mise à jour :PASSAGE AU FRFT 4HV4 GSM Demande de mise à jour
Action:
Passage du forfait 2h vers le 4hv4gsm.
Autre demande d'informations générales : Action:RECU COURRIER RESIL LE
27/07/1998
RESIL A TERME . PAS DE CONSO. JE SUSPEND. ....&
RESILIER.....
Autre demande d'informations générales : Action:RECU COURRIER RESIL DU
28/07/98
Vente terminal neuf suite à action SAV N°INEI remplacé: [] Nouveau N°INEI:
[331004350380060] N°Accord préalable: [] Enseigne-PdV: [750-0850] CLT VENU
4ND B1.12F 9H1TF 1 4VFN1MP 8515 8FC/MA/2/R515 /MTC 1 4G0 764 4ND 8D07R
```

La création dynamique des agrégats



L'intégration et la diffusion des informations



Sommaire

1. *Présentation de Soft Computing*
2. *Positionnement du Data Mining*
3. *Les Apports des Techniques Nouvelles*
4. *Les apports dans l'inter-activité*
5. *Les défis du Data Mining*
- 6. Références**



GESTION DE LA RELATION CLIENT :

La Ménagère de
moins de 50 ans
n'existe plus.

La ménagère de moins de 50 ans n'existe plus : ce n'est plus un marché qui achète, ce sont des Clients, tous différents ; avec leurs habitudes, leurs attentes et leurs besoins spécifiques. Pour les recruter et les fidéliser efficacement, la Gestion de la Relation Client est capitale. Expert depuis plus de 10 ans en CRM, Soft Computing aide à tous niveaux les entreprises à améliorer leur Relation Client.

Services : stratégie e-CRM, datawarehouse marketing, segmentation, scoring, data et webmining, marketing opérationnel, hébergement et enrichissement de bases de données, automatisation du marketing, du service client, des forces de ventes, personnalisation sur Internet et sur les centres d'appels.

En savoir plus : www.softcomputing.com

Soft Computing, 135, rue de la gongre 75116 Paris • Tél. : (33) 01 53 65 61 00

soft.computing

Vous aider à améliorer la Connaissance,
la Personnalisation et la Satisfaction Client.

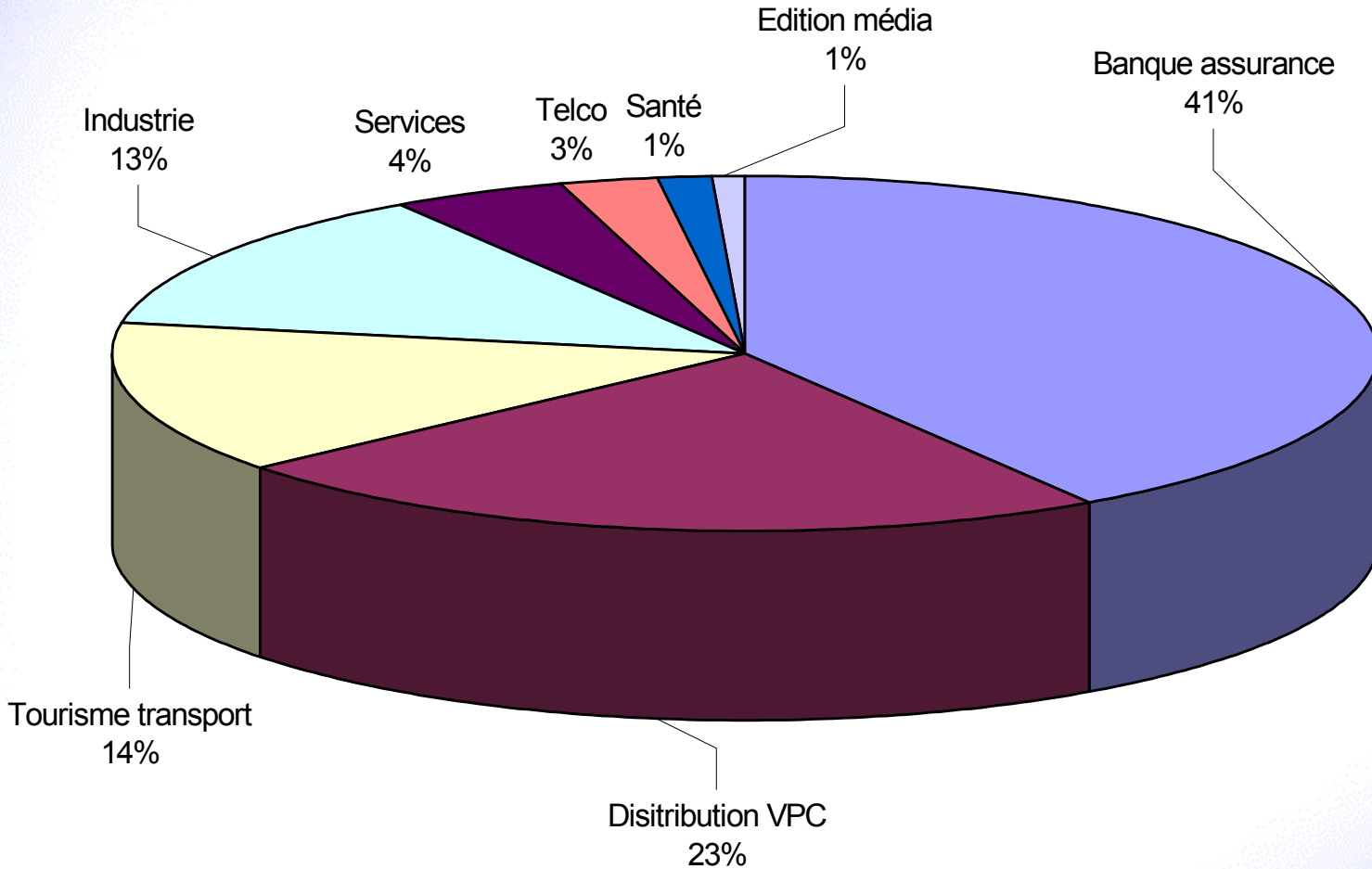
soft.computing

Références pertinentes

Informations commerciales :
efr@sotcomputing.com

46, rue de la Tour
75116 Paris - France
Phone: 01 73 00 55 00
Fax: 01 73 00 55 01
www.softcomputing.com

Des expertises verticales pointues dans tous les secteurs pour lesquels le client compte



Extrait de références : avec tous ceux pour qui le Client compte...

Banque	Assurance	Distribution	VPC	Télécom
Banque Worms, BBL, BNP, Caisse d'Épargne, CCF, CPR Online, Crédit Agricole, Crédit du Nord, Cofinoga, Comdirect, JP-Morgan, Oddo-Pinatton, Selftrade, Société Générale, UBP	AGF, AXA, Autofirst, Azur, DAS, Fortis, Groupama, Lloyd's, La Mondiale, MMA, Eurofil, Okassurance	Agrigel, Casino, Castorama, Club Avantages, Fnac, Match, Mistergoodeal, Printemps, Système U, Promocash, Surcouf, Totalfinal, Yves Rocher	3 Suisses, Aucland, Daxon, Eveil et Jeux, Fnac.com, Guilbert, JPG, France Loisirs, Grand Livre du Mois, IMP, Radiospares, La Redoute, Yves Rocher	9 Telecom, Bouygues Telecom, Debitel, France Telecom, PPR Interactive, Mageos, UUNet
Transport/tourisme	Industrie	PGC	Santé	Média
Air France, Club Med, DHL, Eurodisney, Nouvelles Frontières, Pierre et Vacances, SNCF, UCPA	Alcatel, EDF, Fenwick Linde, GDF, Lex Manutan, Michelin, Provimi Renault, PSA	Petit Bateau, L'oréal, Orcanta, Yves Saint Laurent Parfums, Sony	Smithkline Beecham, Glaxo Wellcome, Ratiopharm	L'alsace, Bertelsmann, EMI Music, Havas Interactive, RTL, TF1, Noos, Europe Régie