



COLLECTION TECHNIQUE ET SCIENTIFIQUE DES TÉLÉCOMMUNICATIONS

Introduction au Data Mining

*Analyse intelligente
des données*

Michel Jambu

 Eyrolles

BIBLIOTHEQUE DU CERIST

Introduction

au

Data Mining

Analyse intelligente des données

BIBLIOTHEQUE DU CERIST

Collection technique et scientifique des télécommunications

Introduction au Data Mining

Analyse intelligente des données

Michel Jambu

*Expert en traitement des données et des connaissances
Direction scientifique - Centre National d'Études des Télécommunications
Professeur et Directeur de Recherches
UFR de Mathématiques de la Décision
Université paris IX - Dauphine*

Deuxième tirage 2000

 **Eyrolles**

IST. 2600

BIBLIOTHEQUE DU CERIST

ÉDITIONS EYROLLES
61, Bld Saint-Germain
75240 Paris Cedex 05
www.eyrolles.com



Le code de la propriété intellectuelle du 1^{er} juillet 1992 interdit en effet expressément la photocopie à usage collectif sans autorisation des ayants droit. Or, cette pratique s'est généralisée notamment dans les établissements d'enseignement, provoquant une baisse brutale des achats de livres, au point que la possibilité même pour les auteurs de créer des œuvres nouvelles et de les faire éditer correctement est aujourd'hui menacée.

En application de la loi du 11 mars 1957, il est interdit de reproduire intégralement ou partiellement le présent ouvrage, sur quelque support que ce soit, sans autorisation de l'Éditeur ou du Centre Français d'exploitation du droit de Copie, 20, rue des Grands Augustins, 75006 Paris.

© Éditions Eyrolles et France Télécom-CNET, 1999, ISBN 2-212-05255-3

Table des Matières

AVANT-PROPOS	V
Chapitre 1. Présentation générale : des données à la décision	1
1. Data Mining, Data Warehousing et analyse des données	1
2. L'analyse des données	1
3. Le Data Warehousing	2
4. Le Data Mining	5
5. Information et décision	6
Chapitre 2. Processus de management de l'information pour la décision	9
1. Data Mining et décision	9
2. Cycle de management de l'information	9
Chapitre 3. Données utilisées en Data Mining	23
1. Données et tableaux de données	23
2. Les variables	24
3. Les individus ou entités	27
4. Les tableaux de données	27
Chapitre 4. Méthodes du Data Mining	43
1. Appréhender des données massives ou complexes	43
2. Sélectionner les données utiles	44
3. Manager et manipuler les données	44
4. Explorer et visualiser les données	45
4.1. Principe	45
4.2. L'analyse exploratoire des données	45
4.3. Les méthodes factorielles	49
5. Segmenter les données	50
5.1. Principe	50

5.2.	<i>La segmentation hiérarchique top-down</i>	52
5.3.	<i>La segmentation hiérarchique bottom-up</i>	53
5.4.	<i>La segmentation-partition</i>	55
6.	Modéliser les données	55
6.1.	<i>Principe</i>	55
6.2.	<i>Les régressions</i>	57
6.3.	<i>La modélisation de séries chronologiques</i>	58
6.4.	<i>La segmentation</i>	59
6.5.	<i>L'analyse discriminante</i>	60
6.6.	<i>La modélisation de règles de décision</i>	60
6.7.	<i>Les méthodes neuronales</i>	60
6.8.	<i>Les méthodes génétiques</i>	61
7.	Décider, prévoir, prédire, pronostiquer	62
7.1.	<i>Principe : les règles de décision</i>	62
7.2.	<i>Les régressions</i>	62
7.3.	<i>La prévision de séries chronologiques</i>	64
7.4.	<i>Les arbres de décision</i>	64
7.5.	<i>Les analyses factorielles</i>	66
7.6.	<i>Les réseaux de neurones</i>	66
7.7.	<i>Les plus proches voisins</i>	67
8.	Les méthodes spécifiques aux données textuelles	68
8.1.	<i>Position du problème</i>	68
8.2.	<i>Constitution de tableaux de données</i>	68
8.3.	<i>Data Mining textuel</i>	69
9.	Modes d'utilisation du Data Mining	69
9.1.	<i>Présentation</i>	69
9.2.	<i>Le mode d'exploration libre</i>	70
9.3.	<i>Le mode d'exploration automatisée</i>	70
9.4.	<i>Le mode d'exploration en temps réel</i>	72
9.5.	<i>Comment "Data Miner"</i>	72
9.6.	<i>Conclusions</i>	72
	Chapitre 5. Evaluation des produits de Data Mining	75
1.	Présentation générale	75
2.	L'information d'entreprise	76
2.1.	<i>Point de vue des décideurs</i>	76
2.2.	<i>Point de vue des directions informatiques</i>	76
2.3.	<i>Point de vue des experts du Data Mining</i>	77
3.	Principe de constitution d'une grille d'évaluation	77

4. Critères et indicateurs d'évaluation	78
Chapitre 6. Problèmes typiques de Data Mining d'entreprise.....	83
1. Data Mining et décideurs	83
2. Les données utiles aux décideurs	84
2.1. Cas général des données d'entreprise	84
2.2. Cas particulier des opérateurs de télécommunications	84
3. Problèmes typiques de Data Mining d'entreprise.....	85
3.1. Etude de cas chez les opérateurs de télécommunications	85
3.2. Améliorer la qualité et l'efficacité du réseau	85
3.3. Améliorer la satisfaction des clients	87
3.4. Améliorer la satisfaction du personnel	89
3.5. Améliorer la relation avec les clients.....	89
3.6. Minimiser l'infidélité des clients	89
3.7. Faire du marketing de niches.....	90
3.8. Faire du marketing local	90
3.9. Augmenter le trafic.....	91
3.10. Prévoir le trafic	91
3.11. Détecter les fraudes	92
3.12. Identifier les tendances de la technologie	92
3.13. Superviser les marchés et la concurrence	93
3.14. Evaluer les appels d'offres	93
3.15. Augmenter la qualité des systèmes d'information	93
3.16. Augmenter la qualité des logiciels	94
3.17. Contrôler les dépenses	94
3.18. Maîtriser les processus de gestion	94
3.19. Visualiser les données de facturation	95
4. Conclusion	95
RÉFÉRENCES BIBLIOGRAPHIQUES	97
RÉPERTOIRE DES PRODUITS DE DATA WAREHOUSING ET DE DATA MINING.....	101
GLOSSAIRE	109