

# Data Mining et statistique décisionnelle: L'intelligence des données

- Author : TUFFERY Stéphane
- Publisher : Editions TECHNIP, 2012
- pages : 848 pages
- N° Class : 621/605



Le data mining et la statistique sont de plus en plus répandus dans les entreprises et les organisations soucieuses d'extraire l'information pertinente de leurs bases de données, qu'elles peuvent utiliser pour expliquer et prévoir les phénomènes qui les concernent (risques, consommation, fidélisation...).

Cette quatrième édition, actualisée et augmentée de 120 pages, fait le point sur le data mining, ses fondements théoriques, ses méthodes, ses outils et ses applications, qui vont du scoring jusqu'au web mining et au text mining. Nombre de ses outils appartiennent à l'analyse des données et la statistique "classique" (analyse factorielle, classification automatique, analyse discriminante, régression logistique, modèles linéaires généralisés, régression pénalisées...) mais certains sont plus spécifiques au data mining, comme les arbres de décision, les réseaux de neurones, les SVM, l'agrégation de modèles et la détection des règles d'associations. Ces outils sont disponibles dans des logiciels de plus en plus puissants et conviviaux, aptes à exécuter de nombreux algorithmes sur de grands volumes de données. Un chapitre de l'ouvrage aide le lecteur à se diriger dans cette offre logicielle et détaille les fonctionnalités des trois principaux logiciels : R, SAS, IBM et SPSS. Ces logiciels sont aussi utilisés pour illustrer par des exemples de nombreuses explications théoriques : une partie de 50 pages est consacrée à une étude de cas complète de credit scoring, qui va de l'exploration des données jusqu'à l'élaboration de la grille de score.

Les aspects méthodologiques vont de la conduite des projets jusqu'aux facteurs de réussite et aux pièges à éviter, en passant par l'évaluation et la comparaison des modèles, leur intégration dans les processus opérationnels, sans oublier les contraintes juridiques dès que l'on traite des données à caractère personnel.

Table des matières :

1. Panorama du data mining.
2. Le déroulement d'une étude de data mining.
3. L'exploration et la préparation des données.
4. L'utilisation des données commerciales et géodémographiques.
5. Les logiciels de statistique et de data mining.
6. Panorama des méthodes de data mining.
7. L'analyse factorielle.
8. Les réseaux de neurones.
9. Les techniques de classification automatique.
10. La recherche des règles d'associations.
11. Les techniques de classement et de prédiction.
12. L'analyse discriminante linéaire et ses généralisations.
13. Le modèle linéaire et ses généralisations.
14. Le modèle logistique et ses généralisations.
15. Les autres modèles prédictifs.
16. L'agrégation de modèles.
17. Une application du data mining : le scoring.
18. Les facteurs de succès d'un projet de data mining.
19. Le text mining.
20. Le web mining.

Annexes. Bibliographie. Index.