

Tufféry, Stéphane

\* 1270.62016

**Data mining and decisional statistics. The intelligence of data. (Data mining et statistique décisionnelle. L'intelligence des données.) 3rd revised and extended ed. (French)**

**Paris: Editions Technip (ISBN 978-2-7108-0946-3). xx, 705 p. EUR 62.00 (2010).**

Publisher's description: 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 troisième édition, actualisée et augmentée de 170 pages, fait le point sur le data mining, 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 (classification automatique, analyse discriminante, régression logistique, modèles linéaires généralisés, régression régularisée...) 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. Un chapitre de l'ouvrage aide d'ailleurs le lecteur à se diriger dans cette offre logicielle et détaille les fonctionnalités des trois principaux: SAS, IBM et SPSS et aussi R, qui bénéficie d'un développement conséquent dans cette troisième édition. Ces logiciels sont aussi utilisés pour illustrer par des exemples de nombreuses explications théoriques: une nouvelle 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.