



# Data mining et statistique décisionnelle - 4ème édition

Stéphane Tufféry

Télécharger

Lire En Ligne

## Data mining et statistique décisionnelle - 4ème édition Stéphane Tufféry

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 à la statistique classique (analyse factorielle, classification automatique, analyse discriminante, régression logistique, modèles linéaires généralisés, régression pénalisée...), 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 et IBM 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 creditscoreing, 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.

[Download Data mining et statistique décisionnelle - 4ème édit ...pdf](#)

[Read Online Data mining et statistique décisionnelle - 4ème éd ...pdf](#)

# **Data mining et statistique décisionnelle - 4ème édition**

*Stéphane Tufféry*

## **Data mining et statistique décisionnelle - 4ème édition Stéphane Tufféry**

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 à la statistique classique (analyse factorielle, classification automatique, analyse discriminante, régression logistique, modèles linéaires généralisés, régression pénalisée...), 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 et IBM 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 creditscoreing, 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.

826 pages

Extrait

Chapitre 1

## Panorama du data mining

Le premier chapitre définit le data mining, et en décrit les principales applications et les apports à différents domaines financiers, industriels, marketing, médicaux et scientifiques. Il situe en outre le data mining par rapport à la statistique, qui lui fournit bon nombre de ses méthodes et concepts théoriques, et par rapport à l'informatique, qui lui fournit sa matière première (les données), ses moyens de calcul et son vecteur de communication (la restitution des résultats) vers les autres applications informatiques et vers les utilisateurs. Il évoque les limites légales inhérentes au traitement de données à caractère personnel, limites érigées pour protéger les libertés individuelles des personnes dont on manipule les données. Ce chapitre s'achève par la mention des principaux facteurs de succès d'un projet de data mining.

### 1.1. Qu'est-ce que le data mining ?

La statistique, naguère confinée dans les études de laboratoire, les expérimentations cliniques, l'actuariat et certaines analyses de risque, envahit aujourd'hui avec le data mining de nombreux domaines, qui vont de l'infini petit (génomique) à l'infiniment grand (astrophysique), du plus quotidien (reconnaissance de l'écriture manuscrite sur les enveloppes) au plus technologique (aide au pilotage aéronautique), du plus ouvert (e-commerce) au plus sécuritaire (prévention du terrorisme, détection de la fraude dans la téléphonie mobile ou l'utilisation des cartes bancaires), du plus industriel (contrôle qualité, pilotage de la production) au plus théorique (enquêtes en sciences humaines, études biologiques, médicales et pharmaceutiques), et du plus alimentaire (études agronomiques et agro-alimentaires) au plus divertissant (prédictions d'audience à la télévision). À cette simple énumération, on devine que le spectre des applications du data mining et de la statistique est très large. Les plus concernés sont les secteurs où d'importants volumes de données doivent être analysés, parfois en vue de prendre des décisions rapides comme le montrent certains des exemples précédents. L'aide à la décision devient une finalité du data mining et de la statistique, dont on n'attend plus seulement (ou plus nécessairement) qu'ils aident à comprendre le réel en le modélisant. Cette démarche n'est pas absolument nouvelle et se rencontre en médecine, où certains traitements ont pu être mis au point sur la base d'analyses statistiques, alors que le mécanisme biologique de la maladie, un cancer par exemple, demeurait mal connu du fait de sa complexité. Le data mining permet de limiter la subjectivité humaine dans les processus de décision, et aussi, grâce à la puissance grandissante des outils informatiques, de traiter de plus en plus rapidement de grands nombres de dossiers.

#### (...) Présentation de l'éditeur

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. Biographie de l'auteur

Stéphane Tufféry est en charge de la statistique dans un grand groupe bancaire. Il enseigne le data mining à l'université Rennes 1 et à l'ENSAI (Ecole nationale de la statistique et de l'analyse de l'information). Il a été traduit en anglais et a récemment publié *Etude de cas en statistique décisionnelle* dans la même collection. Download and Read Online Data mining et statistique décisionnelle - 4ème édition Stéphane Tufféry #B7KLRPJTOFN

Lire Data mining et statistique décisionnelle - 4ème édition par Stéphane Tufféry pour ebook en ligneData mining et statistique décisionnelle - 4ème édition par Stéphane Tufféry Téléchargement gratuit de PDF, livres audio, livres à lire, bons livres à lire, livres bon marché, bons livres, livres en ligne, livres en ligne, revues de livres epub, lecture de livres en ligne, livres à lire en ligne, bibliothèque en ligne, bons livres à lire, PDF Les meilleurs livres à lire, les meilleurs livres pour lire les livres Data mining et statistique décisionnelle - 4ème édition par Stéphane Tufféry à lire en ligne. Online Data mining et statistique décisionnelle - 4ème édition par Stéphane Tufféry ebook Téléchargement PDFData mining et statistique décisionnelle - 4ème édition par Stéphane Tufféry DocData mining et statistique décisionnelle - 4ème édition par Stéphane Tufféry MobipocketData mining et statistique décisionnelle - 4ème édition par Stéphane Tufféry EPub

**B7KLRPJTOFN****B7KLRPJTOFN****B7KLRPJTOFN**