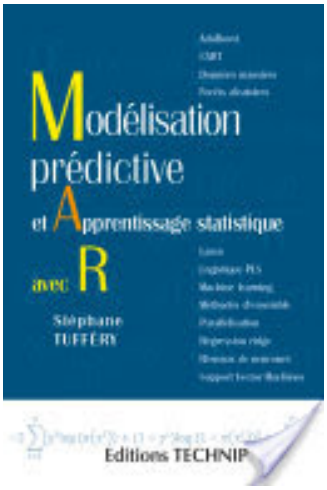

Modélisation prédictive et apprentissage statistique avec R

- **Author : TUFFERY Stéphane**
- **Publisher : Editions TECHNIP, 2015**
- **pages : 432 pages**
- **N° Class : 510/54**

Issu de formations devant des publics variés, cet ouvrage présente les principales méthodes de modélisation de statistique et de machine learning, à travers le fil conducteur d'une étude de cas. Chaque méthode fait l'objet d'un rappel de cours et est accompagnée de références bibliographiques, puis est mise en oeuvre avec des explications détaillées sur les calculs effectués, les interprétations des résultats et jusqu'aux astuces de programmation permettant d'optimiser les temps de calcul. À ce sujet, une annexe est consacrée au traitement des données massives.

L'ouvrage commence par les méthodes de classement classiques les plus éprouvées, mais aborde rapidement les méthodes plus récentes et avancées : régression ridge, lasso, elastic net, boosting, forêts aléatoires, Extra-Trees, réseaux de neurones, séparateurs à vaste marge. Chaque fois, le lien est fait entre la théorie et les résultats obtenus pour montrer qu'ils illustrent bien les principes sous-jacents à ces méthodes. Mais l'aspect pratique est aussi privilégié, avec l'objectif de permettre au lecteur une mise en oeuvre rapide et efficace dans son travail concret. L'exploration et la préparation préliminaire des données sont d'ailleurs décrites, ainsi que le processus de sélection des variables. Une synthèse finale est faite de toutes les méthodes présentées.
La mise en oeuvre s'appuie sur le logiciel libre R et sur un jeu public de données. Ce dernier peut être téléchargé sur internet et présente l'intérêt d'être riche, complet et de permettre des comparaisons grâce aux nombreuses publications dans lesquelles il a servi. Le logiciel statistique utilisé est R, actuellement celui qui se développe le plus : devenu la lingua franca de la statistique et l'outil le plus répandu dans le monde académique, il prend également de plus en plus de place dans le monde de l'entreprise, à tel point que tous les logiciels commerciaux proposent désormais une interface avec
R. Outre qu'il est disponible pour tous, dans de multiples environnements, il est aussi le plus riche statistiquement et c'est le seul logiciel permettant de mettre en oeuvre toutes les méthodes présentées dans cet ouvrage. Enfin, son langage de programmation particulièrement élégant et adapté au calcul athématique permet de se concentrer dans le codage sur les aspects statistiques. R permet d'arriver directement à l'essentiel et de mieux comprendre les méthodes exposées dans l'ouvrage.

Le Code R utilisé dans l'ouvrage est disponible sur cette page dans la partie "Bonus/lire".

Table des matières :