

Big Data et Machine Learning

Les concepts et les outils de la data science



Pirmin Lemberger, Marc Batty
Médéric Morel, Jean-Luc Raffaëlli

Préface d'Aurélien Géron

2^e édition

DUNOD



Pirmin Lemberger, Marc Batty
Médéric Morel, Jean-Luc Raffaëlli

BIG DATA ET MACHINE LEARNING

Manuel du *data scientist*

Cet ouvrage s'adresse à tous ceux qui réfléchissent à la meilleure utilisation possible des données au sein de l'entreprise, qu'ils soient data scientists, DSI, chefs de projets ou spécialistes métier.

Le Big Data s'est imposé comme une innovation majeure pour toutes les entreprises qui cherchent à construire un avantage concurrentiel grâce à l'exploitation de leurs **données** clients, fournisseurs, produits, processus, machines, etc.

Mais quelle solution technique choisir ? Quelles compétences métier développer au sein de la DSI ?

Ce livre est un guide pour comprendre les enjeux d'un projet Big Data, en appréhender les **concepts** sous-jacents (en particulier le **machine learning**) et acquérir les **compétences** nécessaires à la mise en place d'un data lab.

Il combine la présentation :

- de **notions théoriques** (traitement statistique des données, calcul distribué...) ;
- d'**outils** (écosystème Hadoop, Storm...) ;
- d'**exemples de machine learning** ;
- d'une **organisation** typique d'un projet de data science.

Une version de démonstration de « Data Science Studio » et des jeux de données sont téléchargeables à l'adresse suivante : www.dataiku.com/livre-big-data.

MANAGEMENT DES SYSTÈMES D'INFORMATION
APPLICATIONS MÉTIERS
ÉTUDES, DÉVELOPPEMENT, INTÉGRATION
EXPLOITATION ET ADMINISTRATION
RÉSEAUX & TÉLÉCOMS

Préface de
Michel Delattre
DSI du Groupe La Poste

PIRMIN LEMBERGER
Consultant senior et responsable de la veille technologique dans le groupe Sqli, ses tribunes paraissent dans *ITBusiness*, le JDN ainsi que sur le blog du groupe Sqli.

MARC BATTY
Co-fondateur de Dataiku, éditeur de la solution « Data Science Studio », un environnement complet d'analyse et de développement d'applications prédictives.

MÉDÉRIC MOREL
Fondateur et directeur général de Contexo, éditeur de logiciel dans le domaine de la contextualisation.

JEAN-LUC RAFFAËLLI
Directeur de projets stratégiques au sein de la DSI du Groupe La Poste. Il contribue aux différentes déclinaisons SI de la stratégie Business du groupe.



9 782100 720743

5800067
ISBN 978-2-10-072074-3

Les actus

du savoir



Exercice 2 : Echange et Interopérabilité

TAF

- Suivez la démonstration sur le catalogage sur le SIGB PMB ;
- Voir l'utilité des dispositifs d'interopérabilité et d'échange (Z39-50 et Moccam en ligne)
- Z39-50 (Téléchargez la notice de l'auteur Jean-Philippe Accart depuis le serveur PMB de la BNF) ;
- Créer un compte sous : <http://www.moccam-en-ligne.fr> et téléchargez la notice correspondant à l'ISBN : 9783841678430, tout en affichant la description de métadonnées sous le format classique manuel ISBD puis UNIMARC et l'exportez.