

Data Science par Analyse des Données Symboliques :
une nouvelle façon d'analyser les données classiques,
complexes et massives à partir des classes
Applications avec **Syr** et **R**

La numérisation croissante de notre société alimente des bases de données de taille grandissante (Big Data). Ces données sont souvent complexes (hétérogènes et multi-tables) et peuvent être la source de création de valeur considérable à condition qu'elles soient exploitées avec des méthodes d'analyse adéquates. Un « Data Scientist » a justement pour objectif d'extraire des connaissances de ce type de données et c'est l'objectif de cet ouvrage.

Les classes constituent un pivot central de la découverte de connaissances. En Analyse des Données Symboliques (ADS), les classes sont décrites par des variables dites symboliques prenant en compte leur variabilité interne sous forme de distributions, d'intervalles, d'histogrammes, de diagrammes de fréquences, etc.

Le livre débute par la construction de différents types de variables symboliques à partir de classes données. Des statistiques descriptives, une méthode de discrétisation automatique adaptée aux données massives (Big Data) suivies par des indices de proximité étendus aux données symboliques y sont présentés. Vient ensuite un ensemble de méthodes présenté dans le contexte de l'ADS. Il s'agit de la méthode des nuées dynamiques (MND), de la décomposition de mélange par partition (issue de la MND) ou par partition floue (EM), de l'analyse en composantes principales, de l'algorithme Apriori, des règles d'association et des arbres de décision. Pour la prévision, le livre présente des méthodes de régressions dont celles pénalisées « ridge », « lasso » et « elastic », et des séries temporelles.

Pour la mise en application de ces premières méthodes, des exercices et des applications concrètes réalisées auprès d'administrations, d'industriels, de financiers et de scientifiques sont proposés. Leur mise en œuvre s'appuie aussi bien sur le logiciel innovant Syr que sur le logiciel statistique R.

Cet ouvrage d'introduction à l'ADS s'adresse aux étudiants, aux ingénieurs, aux universitaires, ainsi qu'à tous ceux qui désirent comprendre cette nouvelle façon de penser en Science des Données.

Filipe AFONSO est Président de la société SYMBAD spécialisée en Science des Données grâce à des méthodes innovantes développées en interne ou issues de la Recherche universitaire récente. Docteur en Informatique de l'Université Paris Dauphine, il conduit depuis plus de 10 ans des projets d'analyse de données pour le compte des entreprises.

Edwin DIDAY est actuellement Professeur émérite à l'Université Paris-Dauphine. Il est à l'origine de la méthode des nuées dynamiques (ouvrant la voie aux modèles locaux), des pyramides classifiantes (classes avec recouvrement) et de l'Analyse des Données Symboliques (passant de l'analyse des individus à celle des classes). Il est titulaire du prix Montyon décerné par l'Académie des Sciences.

Carole TOQUE, docteur en Statistique de Télécom Paris Tech, poursuit ses recherches sur la prévision des séries temporelles et l'analyse des données à l'Université du Luxembourg. Après plusieurs années d'enseignement en universités et en écoles d'ingénieurs, elle est actuellement statisticienne au Ministère de la Transition écologique et solidaire.

Filipe AFONSO
Edwin DIDAY
Carole TOQUE

Data Science
par Analyse des Données Symboliques

DS Data Science

par Analyse des Données Symboliques

Filipe AFONSO
Edwin DIDAY
Carole TOQUE

Données complexes
Multi-sources
Multi-tables
Big Data

Classes
Description symbolique
Variabilité interne
Discrétisation automatique

Analyse des données
symboliques
Classification
ACP
Prévision

$$-2 \sum_{i=1}^n [y^i \log(\pi(x^i)) + (1 - y^i) \log(1 - \pi(x^i))] + \lambda \sum_{j=1}^p \beta_j^2$$



$$-2 \sum_{i=1}^n [y^i \log(\pi(x^i)) + (1 - y^i) \log(1 - \pi(x^i))] + \lambda \sum_{j=1}^p \beta_j^2$$



$$-2 \sum_{i=1}^n [y^i \log(\pi(x^i)) + (1 - y^i) \log(1 - \pi(x^i))] + \lambda \sum_{j=1}^p \beta_j^2$$

Éditions **TECHNIP**

www.editionstechnip.com