

## Technique de Datamining avec R modélisation et représentation des données



MR-42 3 Jours (21 Heures)



### Description

Le datamining repose sur la maîtrise des techniques d'exploration de données fondamentales : statistiques descriptives, prédictives ou exploratoires. Ce stage pratique vous présentera des méthodes telles que les régressions et les ACP et vous apprendra à les mettre en oeuvre avec le logiciel R.

### À qui s'adresse cette formation ?

#### Pour qui

Responsables Infocentre / Datamining / Marketing / Qualité, utilisateurs et gestionnaires métiers de bases de données.

#### Prérequis

Aucune

### Les objectifs de la formation

- Comprendre le principe de la modélisation statistique Choisir le type de régression en fonction du type de données Réaliser des prédictions Créer des sélections et des classements dans de grands volumes de données pour dégager des tendances

### Programme de la formation

#### Introduction à la modélisation

- Modélisation : la régression.
- Modélisation statistique : rappels des tests statistiques.
- Analyse de données.
- Introduction au logiciel R.
- Travaux pratiques Présentation de plusieurs exemples de modélisation.
- Installation de R et des packages à manipuler.
- Applications sur R des tests et interprétations sur exemples.

#### Analyse de régression linéaire

- Principe de régression linéaire.
- Régression simple, lorsque le modèle comporte un seul paramètre pour des données continues.
- Régression multiple, lorsqu'il y a plus d'un paramètre.
- Autres types de modèles pour données continues.
- Travaux pratiques Application pratique sur R.
- Cas de régression simple et de régression multiple.

### Analyse de régression logistique

- Présentation des différents types de régression logistique.
- Régression logistique Binaire.
- Régression logistique Ordinale.
- Régression logistique Multinomiale.
- Travaux pratiques Application sur R avec des cas pratiques pour des cas de données non continues.
- Traitements sur des données à deux modalités, puis avec des modalités ordinales, puis des modalités nominales.

### Analyse en composantes

- Présentation des différents types d'analyses et sélection.
- Analyse en Composantes Principales (ACP).
- Analyse des Correspondances Multiples (ACM).
- Classification Hiérarchique sur Composantes Principales (CHCP).
- Travaux pratiques Les composantes principales permettent de comprendre la structure de covariance des variables initiales et/ou à créer un plus petit nombre de variables à l'aide de cette structure.
- Applications sur R.

### Analyse factorielle des données

- Comprendre le principe de l'analyse factorielle : résumer la structure des données en un nombre inférieur de dimensions.
- Analyse Factorielle des Correspondances (AFC).
- Analyse Factorielle Multiple (AFM).
- Analyse Factorielle pour Données Mixtes (AFDM).
- Travaux pratiques Exercices d'analyse factorielle sur R.
- Identification des "facteurs" sous-jacents des dimensions associées à une variabilité importante.