

## Programme de formation Formation "Data Mining"

### Les objectifs

- Choisir les méthodes de Data Mining appropriées aux questions posées et à la nature des données
- Expliquer l'intérêt et les conditions d'application des méthodes présentées
- Appliquer, à l'aide du logiciel R, les différentes méthodes
- Interpréter les sorties de chaque méthode

### Public :

Tout public

### Pré-requis

Avoir des connaissances de bases en statistique (moyenne, variance,...) ainsi qu'une pratique des fonctions de base du logiciel R (import de fichiers, manipulation des données et des objets R,...)

### Programme

- Panorama des principales méthodes en Data Mining
- Les statistiques descriptives
- **L'Analyse en Composantes Principales (ACP) et l'Analyse des Correspondances Multiples (ACM)**, méthodes décrivant une population sur un ensemble de variables et définissent des variables synthétiques discriminant de façon optimale les individus constituant cette population.
- **La classification par analyse typologique (partitions directes, CAH)** : formation de classes d'individus, de manière à ce que les individus d'une même classe soient les plus ressemblants possible sur un certain nombre de variables et que les classes construites soient les plus différentes possible
- **Segmentation par arbre (CART et Random Forest)** :
  - les arbres de décision (régression) pour caractériser les classes (variations) d'une variable à expliquer, et de prédire cette variable pour de nouvelles observations
  - la méthode des forêts aléatoires pour hiérarchiser l'importance des variables explicatives

### Evaluation des acquis

Questionnaire

### Méthodes pédagogiques

Exposés et échanges

Manipulations et exercices pratiques sur des études de cas mises en oeuvre avec le logiciel R

**Durée** : 14 heures

### Dates, lieux et cout

- Du 12 mai au 13 mai à Paris (75), 810 euros (coût pédagogique net de TVA)
- Du 17 novembre au 18 novembre à Paris (75), 810 euros (coût pédagogique net de TVA)

**Référence** : RANMU