DATA SCIENCE AVEC R

Formation proposée par BIZOÏ RAZVAN

Le Big Bata Analytics 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 œuvre avec le langage R, RStudio et Shiny.

◷DURÉE
5 Jour(s)
▣MODALITÉS
Présentiel / Distanciel
♙PARTICIPANTS
Jusqu'à 10 participants
★LANGUE
Français
LIEU
France entière
♙CERTIFICATION
Attestation de formation
▦MISE À JOUR
11/04/2023

VOTRE FORMATEUR

BR **BIZOÏ RAZVAN**
Formateur professionnel

TARIFS INDICATIFS

Inter-entreprise	**Pas de dates planifiées**
Intra-entreprise	**Sur devis**
Individuelle	**Sur devis**
Durée	**5 Jour(s)**

Tarifs indicatifs, sous réserve d'analyse du besoin et de devis.

◎Objectifs pédagogiques

✓Comprendre le principe de la modélisation statistique
✓Choisir entre la régression et la classification en fonction du type de données
✓Évaluer les performances prédictives d'un algorithme
✓Créer des sélections et des classements dans de grands volumes de données pour dégager des tendances

♙Public cible et prérequis

Public cible

- Responsables Infocentre (Datamining, Marketing, Qualité...), utilisateurs et gestionnaires métiers de bases de données.

Prérequis

- Connaissances de base en statistiques.
- Connaissances de base en langage R.

▥Programme détaillé

Introduction à la modélisation
- Les types de données dans R.
- Importation-exportation de données.

- Techniques pour tracer des courbes et des graphiques.
- Introduction à R Shiny
- Interface utilisateur avec des exemples d'inputs courants.
- Implémentation côté serveur et dialogue avec l'interface, exemple d'outputs courants.
Travaux pratiques
Prise en main des scripts, Notebooks et applications Shiny.

La collecte et la préparation des données
- Préparation d'un jeu de données
- Nettoyage des données et valeurs manquantes
- Identification des corrélations
- Création de nouvelles variables pour aider à résoudre le problème
Travaux pratiques
Mise en pratique de nettoyage de données et de préparation des variables.

Procédures d'évaluation de modèles
- Préparation d'un jeu de données
- Nettoyage des données et valeurs manquantes
- Identification des corrélations
- Les techniques de ré-échantillonnage en jeu d'apprentissage, de validation et de test.
- Test de représentativité des données d'apprentissage.
- Mesures de performance des modèles prédictifs.
- Matrice de confusion, de coût et la courbe ROC et AUC.
Travaux pratiques
Mise en place d'échantillonnage de jeux de donnes. Effectuer des tests d'évaluations sur plusieurs modèles fournis.

Méthodes d’analyse de données
- Analyse en composantes principales.
- Analyse factorielle des correspondances.
- Analyse des correspondances multiples.
- Analyse factorielle pour données mixtes.
- Factorisation par matrices non négatives
- Latent Dirichlet Allocation
- Classification hiérarchique sur composantes principales.
Travaux pratiques
Mise en œuvre de la diminution du nombre des variables et identification des facteurs sous-jacents des dimensions associées à une variabilité importante.

Les algorithmes non supervisés
- Le clustering hiérarchique.
- Le clustering non hiérarchique.
- Les approches mixtes.
Travaux pratiques
Traitements de clustering non supervisés sur plusieurs jeux de données.

Les algorithmes supervisés
- Le principe de régression linéaire univariée.
- La régression multivariée.
- La régression polynomiale.
- La régression régularisée.
- Le Naive Bayes.
- La régression logistique.
Travaux pratiques
Mise en œuvre des régressions et des classifications sur plusieurs types de données.

Analyse de données textuelles
- Collecte et prétraitement des données textuelles.
- Extraction d'entités primaires, d'entités nommées et résolution référentielle.
- Étiquetage grammatical, analyse syntaxique, analyse sémantique.
- Lemmatisation.
- Représentation vectorielle des textes.
- Pondération TF-IDF.
- Word2Vec.

Travaux pratiques
Explorer le contenu d'une base de textes en utilisant l'analyse sémantique latente.

▣Méthodes pédagogiques & évaluation

Méthodes pédagogiques

- ✓ Apports théoriques et retours d'expérience
- ✓ Exercices pratiques et mises en situation
- ✓ Travaux de groupe et échanges entre pairs
- ✓ Études de cas concrets

Modalités d'évaluation

- ✓ Quiz de validation des acquis
- ✓ Mises en situation évaluées
- ✓ Projet fil rouge
- ✓ Feuille de route post-formation

♿Accessibilité

Une formation accessible à tous

Ce programme peut être adapté pour les personnes en situation de handicap. Contactez-nous pour étudier vos besoins spécifiques.