

Data Science

Data Science avec R (H2O)

3 jours (21 heures) | ★★★★★ 4,6/5 | BIG-H2O | Évaluation qualitative de fin de stage |
Formation délivrée en présentiel ou distanciel ⁽¹⁾

Formations Informatique > Big Data > Data Science



À l'issue de ce stage vous serez capable de :

- Maîtriser l'utilisation de la bibliothèque H2O
- Gérer les modèles de Machine Learning et de Deep Learning avec H2O
- Connaître les cas d'utilisation avec Spark.

Niveau requis

Avoir des connaissances en Machine Learning et R.

Public concerné

Statisticiens, consultants Big Data, data analysts, data scientists.

Cette formation :

- Est animée par un consultant-formateur dont les compétences techniques, professionnelles et pédagogiques ont été validées par des diplômes et/ou testées et approuvées par l'éditeur et/ou par M2i Formation
- Bénéficie d'un suivi de son exécution par une feuille de présence émarginée par demi-journée par les stagiaires et le formateur.

(1) Modalité et moyens pédagogique :

Formation délivrée en présentiel ou distanciel * (e-learning, classe virtuelle, présentiel à distance). Le formateur alterne entre méthodes ** démonstrative, interrogative et active (via des travaux pratiques et/ou des mises en situation). La validation des acquis peut se faire via des études de cas, des quiz et/ou une certification.

Les moyens pédagogiques mis en oeuvre (variables suivant les formations) sont : ordinateurs Mac ou PC (sauf pour les cours de l'offre Management), connexion internet fibre, tableau blanc ou paperboard, vidéoprojecteur ou écran tactile interactif (pour le distanciel). Environnements de formation installés sur les postes de travail ou en ligne. Supports de cours et exercices.

* Nous consulter pour la faisabilité en distanciel. ** Ratio variable selon le cours suivi.

Programme

Jour 1

Machine Learning - Fondamentaux

- Introduction au Machine Learning
- Les modèles du Machine Learning
- Deep Learning
- AutoML

Introduction à H2O

- Présentation d'H2O
- Installation
- Introduction à R

Exemple de travaux pratiques (à titre indicatif)

- *Premier pas*

Les arbres

- Introduction
- Les arbres de décision
- Random Forest
- Random Forest avec H2O
- Gradient Boosting Machine
- Gradient Boosting Machine avec H2O

- Overfitting et Train / Validation / Test
- La validation croisée avec H2O

Exemple de travaux pratiques (à titre indicatif)

- *Cas d'étude d'un arbre de décision avec VC*

Jour 2

Les modèles linéaires et autres

- Introduction
- Régression linéaire
- Régression logistique
- Naïve bayésien

- Hyperparamètre (Grid Search)

Exemple de travaux pratiques (à titre indicatif)

- *Cas d'utilisation avec une régression*

Data manipulation

- Charger et exporter la donnée avec H2O
 - Les sources de données
 - HDFS
 - Remote Files
 - Hive Import
 - BDD JDBC

- Exploration de données
- Manipulation de données
 - Combinaisons
 - GroupBy
 - Merge
 - Data Pre-Processing

Deep Learning H2O

- Les réseaux de neurones
- Deep learning et le Grid Search
- Régression en Deep Learning

Jour 3

H2O : architecture et sécurité

- H2O Stack logiciel
- API REST
- Interaction avec R
- Sécuriser les modèles
- SSL Sécurité

Introduction au Sparkling Water

- Cas d'utilisation nominale
- Construction de modèles
- Le munging de données
- Les processus en stream
- Fonctionnalités
- Les sources de données supportées
- Les formats de données supportés
- Les environnements d'exécutions Spark supportés

L'API H2O

- Démarrage des services H2O
- L'allocation mémoire
- Convertir du H2OFrame au RDD
- Convertir du H2OFrame à un DataFrame
- Convertir du RDD au H2OFrame
- Convertir DataFrame au H2OFrame
- Créer un H2OFrame à partir d'une clé existante
- Le mapping des types entre H2OFrame et Spark DataFrame
- Appeler les algorithmes H2O
- Utiliser Spark Data Sources avec H2OFrame
- Lire à partir du H2OFrame
- Enregistrer dans un H2OFrame
- Charger et sauvegarder les options
- Spécifier le mode d'enregistrement à appliquer

H2O en production

- À propos des POJO et des MOJO
- MOJO Quick Start
- POJO Quick Start
- Exemple de design patterns
- Ressources supplémentaires