

Machine learning , l'état de l'art

Référence : **PYDS010**

Durée : **2 jours (14 heures)**

Certification : **Aucune**

Connaissances préalables

- Connaissances des principes du BigData, et des architectures techniques mises en oeuvre

Profil des stagiaires

- Chefs de projet
- Développeurs
- Data scientists
- Architectes

Objectifs

- Savoir définir les étapes de préparation des données, comprendre et mettre en oeuvre l'apprentissage automatique, les techniques de classification de données, les apports des réseaux de neurones et du Deep Learning

Certification préparée

- Aucune

Méthodes pédagogiques

- 6 à 12 personnes maximum par cours, 1 poste de travail par stagiaire
- Remise d'une documentation pédagogique papier ou numérique pendant le stage
- La formation est constituée d'apports théoriques, d'exercices pratiques et de réflexions

Formateur·rice

- Consultant-Formateur expert Bigdata

Méthodes d'évaluation des acquis

- Auto-évaluation des acquis par le stagiaire via un questionnaire
- Attestation des compétences acquises envoyée au stagiaire
- Attestation de fin de stage adressée avec la facture

Contenu du cours

1. Introduction

- Zoom sur les données : format, volumes, structures, ... et les requêtes, attentes des utilisateurs
- Etapes de la préparation des données
- Définitions, présentation du data munging
- Le rôle du data scientist

2. Gouvernance des données

- Qualité des données
- Transformation de l'information en donnée
- Qualification et enrichissement
- Sécurisation et étanchéité des lacs de données
- Flux de données et organisation dans l'entreprise
- De la donnée maître à la donnée de travail
- MDM
- Mise en oeuvre pratique des différentes phases : nettoyage, enrichissement, organisation des données

3. Traitements statistiques de base

- Introduction aux calculs statistiques
- Paramétrisation des fonctions
- Applications aux fermes de calculs distribués
- Problématiques induites
- Approximations
- Précision des estimations

4. Data Mining

- Besoin, apports et enjeux
- Extraction et organisation des classes de données
- Analyse factorielle

5. Machine Learning

- Apprentissage automatique
- Définition, les attentes par rapport au Machine Learning
- Les valeurs d'observation, et les variables cibles
- Ingénierie des variables
- Les méthodes : apprentissage supervisé et non supervisé
- Classification des données
- Algorithmes : régression linéaire, k-moyennes, k-voisins, classification naïve bayésienne, arbres de décision, forêts aléatoires, etc.
- Création de jeux d'essai, entraînement et construction de modèles
- Prévisions à partir de données réelles
- Mesure de l'efficacité des algorithmes
- Courbes ROC.
- Parallélisation des algorithmes
- Choix automatique

6. IA

- Introduction aux réseaux de neurones
- Réseaux de neurones à convolution
- Modèles de CNN
- Les types de couches : convolution, pooling et pertes
- L'approche du Deep Learning
- DeepLearning4j sur Spark

7. Les risques et écueils

- Importance de la préparation des données
- L'écueil du "surapprentissage"

8. Visualisation des données

- L'intérêt de la visualisation
- Outils disponibles
- Exemples de visualisation avec R et Python

Notre référent handicap se tient à votre disposition au [01.71.19.70.30](tel:0171197030) ou par mail à referent.handicap@edugroupe.com pour recueillir vos éventuels besoins d'aménagements, afin de vous offrir la meilleure expérience possible.