

BEST *Data Science : niveau avancé*

DESCRIPTION

Le Big Data et la data science sont en train de changer le monde que nous connaissons. Avec la digitalisation sans cesse croissante de nos interactions et les progrès accomplis en matière d'intelligence artificielle, les entreprises ont de plus en plus recours aux algorithmes pour améliorer leurs processus de prise de décision et proposer de nouveaux services. Le Data scientist, à travers sa faculté à trouver à partir de l'exploitation des données une réponse adéquate à un problème fonctionnel donné, est au cœur de ces évolutions. Après le module sur les bases de la Data science, la formation « Data Science : niveau avancé » vous permettra d'approfondir certains des concepts avancés de machine learning (feature engineering avancé, modèles ensemblistes...) utilisés notamment par les lauréats de concours Kaggle. Vous découvrirez également les techniques d'analyse et de traitement spécifiques aux données de type image ou texte (deep learning, réseaux de neurones convolutifs, réseaux récurrents,...).

OBJECTIFS PÉDAGOGIQUES

Découvrir et utiliser la puissance prédictive des modèles ensemblistes
Savoir effectuer un "feature engineering" performant
Appréhender les techniques de text-mining et de deep-learning à travers des exemples concrets
Enrichir sa boîte à outils de data scientist

PUBLIC CIBLE

Analyste
Statisticien
Architecte
Développeur
Data scientist

PRÉ-REQUIS

- Connaissances de base en programmation ou en scripting
- Avoir suivi la formation "Fondamentaux de la Data Science" (DSDFX) serait en plus

MÉTHODE PÉDAGOGIQUE

Formation avec apports théoriques, échanges sur les contextes des participants et retours d'expérience pratique du formateur, complétés de travaux pratiques et de mises en situation.

PROFILS DES INTERVENANTS

Toutes nos formations sont animées par des consultants-formateurs expérimentés et reconnus par leurs pairs.

MODALITÉS D'ÉVALUATION ET FORMALISATION À L'ISSUE DE LA FORMATION

L'évaluation des acquis se fait tout au long de la session au travers des

Stage pratique en présentiel DATA SCIENCE

Code :
DSNVA

Durée :
3 jours (21 heures)

Exposés :
50%

Cas pratiques :
40%

Échanges d'expérience :
10%

Sessions à venir :

10 - 12 fév. 2021
Formation à distance / 2 400 eur

10 - 12 mai 2021
Formation à distance / 2 400 eur

19 - 21 juil. 2021
Paris / 2 400 eur

27 - 29 sept. 2021
Paris / 2 400 eur

24 - 26 nov. 2021
Paris / 2 400 eur

Tarif & dates intra :
Sur demande

ateliers et des mises en pratique. Une évaluation à chaud sur la satisfaction des stagiaires est réalisée systématiquement en fin de session et une attestation de formation est délivrée aux participants mentionnant les objectifs de la formation, la nature, le programme et la durée de l'action de formation ainsi que la formalisation des acquis.

POUR ALLER PLUS LOIN :

- Formation "Fondamentaux de la Data Science" (DSFDX)
- Ouvrage "Data Science : Fondamentaux et études de cas" - Eric Biernat, Michel Lutz
- Ouvrage "The Elements of Statistical Learning: Data Mining, Inference, and Prediction" - Hastie, Tibshirani, Friedman
- Toutes nos formations Data Science
- Nos formations Hadoop
- Nos formations NoSQL

Programme pédagogique détaillé par journée

Jour 1

RAPPEL DES FONDAMENTAUX

- Ecosystème Big Data et Data Science
- Comment modéliser un problème de data science ?
- Les différentes familles d'algorithmes (supervisé : classification/régression, non supervisé)
- Les algorithmes classiques
- Comment évaluer la performance ?
- Sur apprentissage et compromis biais/variance

MODÈLES ENSEMBLISTES

- Rappels
- Pourquoi ça fonctionne ? Raisons théoriques
- Introduction au stacking
 - Architecture et promesses du stacking
 - Feature weighted stacking
 - Mise en application

INTRODUCTION AU TEXT MINING

- Un modèle de représentation : le bag of words
- Normalisations usuelles
- Stemming, lemmatization
- Distances (Levenshtein, Hamming, Jaro-Winkler)
- Word2Vec

Jour 2

FEATURE ENGINEERING AVANCÉ

- Normalisation
 - Qu'est ce que la normalisation ?
 - Quand l'utiliser ?
- Réduction de dimension (ACP, TSNE, LSA, etc.)
- Transformation et interactions entre variables
- Traitement des variables catégorielles à haute dimensionnalité
- Création de variables extraites d'arbres (Facebook Trick)

RÉSEAUX DE NEURONES ET DEEP LEARNING

- L'origine : le perceptron
- Les réseaux de neurones
- Deep learning
 - Objectif : s'affranchir du feature engineering manuel
 - Convolution
 - Réseaux récurrents
- Cas concret : reconnaissance de chiffres

APPRENTISSAGE SEMI-SUPERVISÉ

Jour 3

RAPPELS ET RÉVISIONS

- Synthèse des points abordés en journées 1 et 2
- Approfondissement des sujets sélectionnés avec l'intervenant

MISE EN PRATIQUE

- Le dernier jour est entièrement consacré à des mises en pratique

SÉLECTION ET PARTICIPATION À UNE COMPÉTITION

- Le formateur sélectionnera une compétition en cours sur Kaggle qui sera démarrée en jour 3 par l'ensemble des participants