Data Science : niveau avancé

Approfondir des concepts avancés de machine learning et enrichir sa boîte à outils de Data Scientist

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 PEDAGOGIQUES

- Découvrir et utiliser la puissance prédictive des modèles ensemblistes
- Effectuer un "feature engineering" performant
- Préciser les techniques de text-mining et de deep-learning à travers des exemples concrets
- Compléter sa boite à outils de data scientist

PUBLIC CIBLE

- Analyste
- Statisticien
- Architecte
- Développeur
- Data scientist

PRE-REQUIS

- Connaissances de base en programmation (python notamment) ou en scripting
- Avoir suivi la formation "Fondamentaux de la Data Science"

Stage pratique
Data Science

Code :
DSNVA

Durée :
3 jour(s) (21,00 heures)

Exposés : **50 %**
Cas pratiques : **40 %**
Echanges d'expérience : **10 %**

Inter-entreprises :
Prochaines sessions disponibles sur notre site web.
Tarif : 2 510,00 € HT / participant

Intra-entreprise :
Tarifs et dates sur demande.

(DSDFX) serait en plus

METHODE PEDAGOGIQUE
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.

PROFIL DES INTERVENANTS

Cette formation est dispensée par un·e ou plusieurs consultant·es d'OCTO Technology ou de son réseau de partenaires, expert·es reconnus des sujets traités.

Le processus de sélection de nos formateurs et formatrices est exigeant et repose sur une évaluation rigoureuse leurs capacités techniques, de leur expérience professionnelle et de leurs compétences pédagogiques.

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 ateliers et des mises en pratique.

Afin de valider les compétences acquises lors de la formation, un formulaire d'auto-positionnement est envoyé en amont et en aval de celle-ci.

En l'absence de réponse d'un ou plusieurs participants, un temps sera consacré en ouverture de session pour prendre connaissance du positionnement de chaque stagiaire sur les objectifs pédagogiques évalués.

Une évaluation à chaud est également effectuée en fin de session pour mesurer la satisfaction des stagiaires et un certificat de réalisation leur est adressé individuellement.

PROGRAMME PEDAGOGIQUE DETAILLE

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

Accessibilité

L'inclusion est sujet important pour OCTO Academy.
Nos référent·es sont à votre disposition pour faciliter l'adaptation de votre formation à vos besoins spécifiques.

Pour les contacter : academy.accessibilite@octo.com