MACHINE LEARNING LES BASES DE L'APPRENTISSAGE MACHINE

Public	Ingénieurs, analystes, responsables marketing, Data Analysts, Data Scientists, Data Steward.
Durée	3 jours - 21 heures
Pré-requis	Connaître l'utilité du Data Mining et les problématiques du Big Data dans le ciblage économique
Objectifs	Comprendre les différences entre apprentissage automatique supervisé, non supervisé et méta-apprentissage Savoir transformer un gros volume de données à priori hétérogènes en informations utiles Maîtriser l'utilisation d'algorithmes d'auto-apprentissage adaptés à une solution d'analyse Comprendre comment exploiter de gros volumes de données textuelles Être capable d'appliquer ces différentes techniques aux projets Big Data
Méthodes pédagogiques	Pour bien préparer la formation, le stagiaire remplit une évaluation de positionnement et fixe ses objectifs à travers un questionnaire. La formation est délivrée en présentiel ou distanciel (e-learning, classe virtuelle, présentiel et à distance). Le formateur alterne entre méthodes démonstratives, interrogatives et actives (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. 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 Audit Conseil Formation.
Moyens techniques	1 poste de travail complet par personne De nombreux exercices d'application Mise en place d'ateliers pratiques Remise d'un support de cours Passage de certification(s) dans le cadre du CPF Remise d'une attestation de stage
Modalité d'évaluation des acquis	Evaluation des besoins et objectifs en pré et post formation Evaluation technique des connaissances en pré et post formation Evaluation générale du stage
Délai d'accès	L'inscription à cette formation est possible jusqu'à 5 jours ouvrés avant le début de la session
Accessibilité handicapés	Au centre d'affaires ELITE partenaire d'ACF à 20 m. Guide d'accessibilité à l'accueil.

L'APPRENTISSAGE MACHINE

- Introduction
- Champs de compétences
- Focus Data Science (Data Mining)
- Focus Machine Learning
- Focus Big Data
- Focus Deep Learning
- Définition de l'apprentissage machine
- Exemples de tâches du machine Learning
- Que peuvent apprendre les machines
- Les différents modes d'entraînement

LES FONDAMENTAUX DE L'APPRENTISSAGE MACHINE

- Préambule : - Un problème d'optimisation - Quête de la capacité optimale du modèle - Relation capacité et erreurs - Un apport philosophique - Cadre statistique - Anatomie d'un modèle d'apprentissage machine
- Jeux de données d'entraînement : - Cadre statistique - Les variables prédictives - Chaîne de traitement des variables prédictives - Les variables à prédire
- Fonctions hypothèses : - Principe : jeux de fonctions hypothèses - Contexte de sélection des fonctions hypothèses - Caractéristiques des fonctions hypothèses - Modèles probabilistes Fréquentistes et Bayésiens
- Fonctions de coûts : - Les estimateurs - Principe du maximum de vraisemblance (MLE*) - MAP - Maximum A Posteriori - Le biais d'un estimateur - La variance d'un estimateur - Le compromis biais - variance - Les fonctions de coûts - La régularisation des paramètres
- Algorithmes d'optimisations : - Les grandes classes d'algorithmes d'optimisation - La descente de gradient (1er ordre) - Descente de gradient (détails) - Les approches de Newton (2nd ordre) - Optimisation batch et stochastique - Pour aller plus loin
- Lab : Mise en oeuvre de l'environnement de travail machine Learning

LA CLASSIFICATION

- Introduction : - Choisir un algorithme de classification
- La régression logistique : - Du Perceptron à la régression logistique - Hypothèses du modèle - Apprentissage des poids du modèle - Exemple d'implémentation : scikit-learn - Régression logistique - Fiche Synthèse
- SVM : - Classification à marge maximum - La notion de marge souple (soft margin) - Les machines à noyau (kernel machines) - L'astuce du noyau (kernel trick) - Les fonctions noyaux - SVM - Maths - SVM - Fiche Synthèse
- Arbres de décision : - Principe de base - Fonctionnement - Maximisation du Gain Informationnel - Mesure d'impureté d'un noeud - Exemple d'implémentation : scikit-learn -Arbres de décision - Fiche Synthèse
- K plus proches voisins (kNN) : - L'apprentissage à base d'exemples - Principe de fonctionnement - Avantages et désavantages - kNN - Fiche synthèse
- Synthèse
- Lab : Expérimentation des algorithmes de classification sur cas concrets

LES PRATIQUES

- Prétraitement : - Gestion des données manquantes - Transformateurs et estimateurs - Le traitement des données catégorielles - Le partitionnement des jeux de données - Mise à l'échelle des données
- Ingénierie des variables prédictives (Feature Engineering) : - Sélection des variables prédictives - Sélection induite par régularisation L1 - Sélection séquentielle des variables - Déterminer l'importance des variables - Réduction dimensionnelle par Compression des données - L'extraction de variables prédictives - Analyse en composante principale (ACP) - Analyse linéaire discriminante (ADL) - l'ACP à noyau (KPCA)
- Réglages des hyper-paramètres et évaluation des modèles : - Bonnes pratiques - La notion de Pipeline - La validation croisée (cross validation) - Courbes d'apprentissage - Courbes de validation - La recherche par grille (grid search) - Validation croisée imbriquée (grid searchcv) - Métriques de performance
- Synthèse
- Lab : Expérimentation des pratiques du machine learning sur cas concrets

L'APPRENTISSAGE D'ENSEMBLES (ENSEMBLE LEARNING)

- Introduction
- L'approche par vote
- Une variante : l'empilement (stacking)
- Le bagging
- Les forêts aléatoires
- Le boosting
- La variante Adaboost
- Gradient Boosting
- Fiches synthèses
- Lab : L'apprentissage d'ensemble sur un cas concret

LA RÉGRESSION

- Régression linéaire simple
- Régression linéaire multi-variée
- Relations entre les variables
- Valeurs aberrantes (RANSAC)
- Évaluation de la performance des modèles de régression
- La régularisation des modèles de régression linéaire
- Régression polynomiale
- La régression avec les forêts aléatoires
- Synthèse
- Lab : La régression sur un cas concret

LE CLUSTERING

- Introduction
- Le regroupement d'objets par similarité avec les k-moyens (k-means)
- k-means : algorithme
- L'inertie d'un cluster
- Variante k-means ++
- Le clustering flou
- Trouver le nombre optimal de clusters avec la méthode Elbow
- Appréhender la qualité des clusters avec la méthode des silhouettes
- Le clustering hiérarchique
- Le clustering par mesure de densité DBSCAN
- Autres approches du Clustering
- Synthèse
- Lab : Le clustering sur un cas concret

NOUS CONTACTER

Siège social
16, ALLÉE FRANÇOIS VILLON
38130 ÉCHIROLLES

Téléphone
04 76 23 20 50 - 06 81 73 19 35

Suivez-nous sur les réseaux sociaux, rejoignez la communauté !

ACF Audit Conseil Formation

@ACF_Formation

Dernière mise à jour : 05/09/2024

PROFIL Formateur : Les formateurs sont recrutés selon plusieurs critères :
Expérience, pédagogie, dynamisme et prévoyance.