



ORGANISME DE FORMATION AUX TECHNOLOGIES ET METIERS DE L'INFORMATIQUE

Formation Les bases de l'apprentissage Machine (Machine Learning)

Transformer des volumes massifs de données en informations utiles

N° ACTIVITÉ : 11 92 18558 92

TÉLÉPHONE : 01 85 77 07 07

E-MAIL : inscription@hubformation.com

La maîtrise du Data Mining et du Machine Learning est devenue une compétence nécessaire, voire même indispensable à toute personne souhaitant développer une expertise Big Data puisqu'elle permet d'explorer ou de fouiller de très importants volumes de données pour construire des modèles et répondre aux problèmes très variés des entreprises lorsque les méthodes statistiques traditionnelles deviennent inopérantes. Pour cela, les experts en Big Data doivent maîtriser l'élaboration et l'étude des algorithmes permettant à des machines d'apprendre automatiquement à partir des données et à effectuer des tâches de façon autonome pour modéliser des tendances.

Objectifs

- | Identifier les différences entre apprentissage automatique supervisé, non supervisé et méta-apprentissage
- | Transformer un gros volume de données à priori hétérogènes en informations utiles
- | Utiliser les algorithmes d'auto-apprentissage adaptés à une solution d'analyse
- | Exploiter de gros volumes de données textuelles
- | Appliquer ces différentes techniques aux projets Big Data

Public

- | Ingénieurs, analystes, responsables marketing
- | Data Analysts, Data Scientists, Data Steward
- | Toute personne intéressée par les techniques de Data Mining et de Machine Learning

Prérequis

- | Connaître l'utilité du Data Mining et les problématiques du Big Data dans le ciblage économique

Programme de la formation

L'apprentissage machine (Introduction)

- | 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

- | Un problème d'optimisation
- | Quête de la capacité optimale du modèle
- | Relation capacité et erreurs

Référence	DEC105
Durée	3 jours (21h)
Tarif	2 350 €HT

SESSIONS PROGRAMMÉES

A DISTANCE (FRA)

- du 10 au 12 juin 2025
- du 28 au 30 juillet 2025

PARIS

- du 10 au 12 juin 2025
- du 28 au 30 juillet 2025

AIX-EN-PROVENCE

- du 10 au 12 juin 2025
- du 1er au 3 septembre 2025

BORDEAUX

- du 10 au 12 juin 2025
- du 28 au 30 juillet 2025

GRENOBLE

- du 10 au 12 juin 2025
- du 28 au 30 juillet 2025

LILLE

- du 28 au 30 juillet 2025
- du 1er au 3 septembre 2025

LYON

- du 10 au 12 juin 2025
- du 28 au 30 juillet 2025

NANTES

- du 28 au 30 juillet 2025
- du 20 au 22 octobre 2025

[VOIR TOUTES LES DATES](#)

- | 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
- | 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
- | 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

Méthode pédagogique

Chaque participant travaille sur un poste informatique qui lui est dédié. Un support de cours lui est remis soit en début soit en fin de cours. La théorie est complétée par des cas pratiques ou exercices corrigés et discutés avec le formateur. Le formateur projette une présentation pour animer la formation et reste disponible pour répondre à toutes les questions.

Méthode d'évaluation

Tout au long de la formation, les exercices et mises en situation permettent de valider et contrôler les acquis du stagiaire. En fin de formation, le stagiaire complète un QCM d'auto-évaluation.

Suivre cette formation à distance

Voici les prérequis techniques pour pouvoir suivre le cours à distance :

- | Un ordinateur avec webcam, micro, haut-parleur et un navigateur (de préférence Chrome ou Firefox). Un casque n'est pas nécessaire suivant l'environnement.
 - | Une connexion Internet de type ADSL ou supérieure. Attention, une connexion Internet ne permettant pas, par exemple, de recevoir la télévision par Internet, ne sera pas suffisante, cela engendrera des déconnexions intempestives du stagiaire et dérangera toute la classe.
 - | Privilégier une connexion filaire plutôt que le Wifi.
 - | Avoir accès au poste depuis lequel vous suivrez le cours à distance au moins 2 jours avant la formation pour effectuer les tests de connexion préalables.
 - | Votre numéro de téléphone portable (pour l'envoi du mot de passe d'accès aux supports de cours et pour une messagerie instantanée autre que celle intégrée à la classe virtuelle).
 - | Selon la formation, une configuration spécifique de votre machine peut être attendue, merci de nous contacter.
 - | Pour les formations incluant le passage d'une certification la dernière journée, un voucher vous est fourni pour passer l'examen en ligne.
 - | Pour les formations logiciel (Adobe, Microsoft Office...), il est nécessaire d'avoir le logiciel installé sur votre machine, nous ne fournissons pas de licence ou de version test.
 - | Horaires identiques au présentiel.
-

Accessibilité



Les sessions de formation se déroulent sur des sites différents selon les villes ou les dates, merci de nous contacter pour vérifier l'accessibilité aux personnes à mobilité réduite.
Pour tout besoin spécifique (vue, audition...), veuillez nous contacter au 01 85 77 07 07.