

Machine Learning - Fondamentaux

MACHLEARN

La maîtrise du Data Mining et du Machine Learning est aujourd'hui essentielle pour développer une expertise en Big Data. Ces outils permettent d'analyser de très grands volumes de données et de construire des modèles adaptés aux besoins des entreprises lorsque les méthodes classiques ne suffisent plus. Les experts doivent ainsi savoir concevoir des algorithmes capables d'apprendre automatiquement à partir des données pour identifier et modéliser les tendances.

Présentiel - Synchrones

Public Visé

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

Pré Requis

Comprendre l'utilité du Data Mining et les problématiques du Big Data dans le ciblage économique

Objectifs pédagogiques et d'évaluation

À l'issue de la formation, le participant sera en mesure de :

- Identifier les différences entre l'apprentissage supervisé, non supervisé et le méta-apprentissage
- Savoir transformer de grands volumes de données hétérogènes en informations exploitables
- Maîtriser l'utilisation d'algorithmes d'apprentissage automatique adaptés aux besoins d'analyse
- Comprendre les méthodes permettant d'exploiter de larges ensembles de données textuelles
- Etre capable de mettre en œuvre ces techniques dans des projets Big Data

Méthodes pédagogiques

- Beaucoup de tests et d'exemples concrets en lien direct avec les besoins des participants
- Des techniques directement mobilisables dans le cadre de leurs fonctions
- Formation axée sur la mise en pratique
- Méthode participative
- Mise en situation professionnelle
- Alternance de cours et d'exercices dirigés ou en autonomie
- Travaux individuels et corrections collectives
- Evaluation croisée et partage de bonnes pratiques
- Quizz

Parcours pédagogique

1 — Introduction au Machine Learning

- Présentation générale
- Domaines d'application
- Focus Data Science (Data Mining)
- Focus Machine Learning
- Focus Big Data
- Focus Deep Learning
- Définition du Machine Learning
- Exemples de tâches traitées par le Machine Learning
- Types d'apprentissages
- Ce que les machines peuvent apprendre
- Les différents modes d'entraînement

2 — Les fondamentaux du Machine Learning

- Notions clés
- Problème d'optimisation
- Capacité optimale du modèle
- Lien entre capacité et erreurs
- Cadre statistique
- Anatomie d'un modèle de Machine Learning
- Jeux de données d'apprentissage
- Variables prédictives
- Chaîne de traitement des variables
- Variables cibles
- Fonctions hypothèses
- Principe et rôle
- Sélection et caractéristiques
- Approches fréquentistes et bayésiennes
- Fonctions de coût et estimation
- Estimateurs
- Maximum de vraisemblance (MLE)
- Maximum a posteriori (MAP)
- Biais et variance
- Compromis biais-variance
- Régularisation

Algorithmes d'optimisation
Grandes familles d'algorithmes
Descente de gradient
Méthodes de Newton
Optimisation batch et stochastique

3 — La classification

Introduction
Choisir un algorithme de classification
Régression logistique
Du perceptron à la régression logistique
Hypothèses et apprentissage des paramètres
Exemple avec scikit-learn
SVM (Support Vector Machines)
Principe de marge maximale
Marges souples
Méthodes à noyaux et kernel trick
Arbres de décision
Principe et fonctionnement
Gain informationnel
Mesures d'impureté
Exemple avec scikit-learn
k-plus proches voisins (kNN)
Principe
Points forts et limites
Synthèse des méthodes

4 — Les bonnes pratiques en Machine Learning

Prétraitement des données
Gestion des données manquantes
Traitement des variables catégorielles
Partitionnement des jeux de données
Mise à l'échelle
Feature Engineering
Sélection de variables
Régularisation L1
Sélection séquentielle
Importance des variables
Réduction dimensionnelle
ACP, LDA, ACP à noyau
Réglage et évaluation des modèles
Notion de pipeline
Validation croisée
Courbes d'apprentissage et de validation
Grid search
Validation croisée imbriquée
Métriques de performance

5 — L'apprentissage par ensembles (Ensemble Learning)

Principe du vote
Empilement (stacking)
Bagging
Forêts aléatoires
Boosting
AdaBoost
Gradient Boosting

6 — La régression

Régression linéaire simple et multiple
Analyse des relations entre variables
Détection des valeurs aberrantes (RANSAC)
Évaluation des performances
Régularisation
Régression polynomiale

Régression avec forêts aléatoires
Synthèse

7 — Le clustering

Introduction au clustering
Méthode k-means et variantes
Notion d'inertie
k-means++
Clustering flou
Choix du nombre de clusters (méthode Elbow)
Évaluation par silhouettes
Clustering hiérarchique
DBSCAN
Autres approches de clustering
Synthèse

Moyens pédagogiques

1 ordinateur par stagiaire - Salle de formation claire, climatisée et spacieuse - Tableau blanc - Vidéo projecteur - Support de cours - Logiciel d'assistance des stagiaires à distance

Qualification Intervenant-e-s

Data Scientist - 20 ans d'expérience en analyse de données au sein d'équipes CRM et Marketing - 5 ans d'enseignement du module Datamining à l'université de Lille



Méthodes et modalités d'évaluation

Evaluation diagnostique en amont : questionnaire d'auto-positionnement et recueil des attentes & besoins - Evaluation formative en cours de formation : exercices - Evaluation sommative en fin de formation : exercice/QCM - Bilan individuel des compétences acquises – Attestation de fin de formation - Questionnaire de satisfaction à chaud - Questionnaire de satisfaction à froid envoyé 15 jours après la formation pour assurer le suivi post formation

Modalités d'accessibilité handicap

Nos formations sont accessibles aux personnes en situation de handicap et aux besoins spécifiques. Une étude personnalisée avec notre référente handicap, permettra d'adapter les moyens pédagogiques, techniques ou organisationnels.



Durée

21.00 Heures

3

Jours

Effectif

De 3 à 8 Personnes