

Formation Machine Learning

Mise à jour janvier 2025

Inter 1800€ HT/participant

Intra 6750€ HT* groupe de 6 participants

*hors frais de déplacement et de personnalisation de formation sur-mesure

Cette formation vous permettra d'acquérir les bases du Machine Learning. À la fin de ces trois jours serez en mesure de créer et de déployer vos propres modèles prédictifs, à l'aide de l'écosystème scientifique du langage Python.

Vous aurez alors à votre disposition tout un ensemble de bibliothèques et d'outils open-source, tels que Numpy, Scikit-Learn ou TensorFlow.

Les concepts théoriques étudiés seront accompagnés d'exemples pratiques, utilisant des données issues de problèmes réels et de domaines variés (biologie, analyse d'images, climatologie, etc).

Durée: 21.00 heures (3.00 jours)

À QUI S'ADRESSE CETTE FORMATION ?

Profil du participant

- Ingénieurs, Développeurs, Chercheurs
- Data scientist

Prérequis

- Connaissances de base en algorithmie
- Connaissances de base en mathématiques (probabilités et statistiques : notion de variable aléatoire, de probabilité conditionnelle, de moyenne et de variance / vecteurs et espaces de dimensions N et fonction à N variables)
- Connaissance du langage Python (structures algorithmiques de base (if, for, while) / chaînes, booléens, nombres flottants et entiers / listes, tuples, dictionnaires / les fonctions / visibilité des variables / utilisation des packages et modules)

OBJECTIFS PÉDAGOGIQUES

1. Comprendre en quoi consiste l'apprentissage automatique
2. Identifier les problèmes pour lesquels ce type de méthode saura apporter une réponse pertinente
3. Connaître les différents types de méthodes et savoir utiliser les principaux algorithmes

CONTENU (PROGRESSION PÉDAGOGIQUE)

Les bases de l'apprentissage automatique

- Qu'est ce que l'apprentissage automatique ?
- Apprentissage supervisé
- Apprentissage non-supervisé
- Apprentissage par renforcement
- Présentation de quelques applications remarquables

La boîte à outil du data scientist

- Anaconda, l'une des distributions Python scientifique les plus populaires
 - TP : Présentation des Jupyter Notebooks et de iPython
- Numpy, une bibliothèque Python de calcul numérique (algèbre linéaire) permettant de manipuler des vecteurs et matrices à plusieurs dimensions
 - TP : Numpy : Analyse statistique rapide de la population française : création de tableaux, accès aux données, analyse des données, lecture de fichier
- Pandas, une bibliothèque permettant d'analyser avec efficacité des structures de données complexe
 - TP : Analyse statistique de données œnologiques : création de Série et de DataFrame, accès aux données, agrégation des données, analyse statistique
- Matplotlib, pour nous permettre de générer quelques visualisations.
 - TP : Visualisations autour de la démographie française

Manipulation des données

- Quelles données utiliser ?
 - TP : Prédire la réussite d'un Kickstarter
- Constitution d'un ensemble de données d'apprentissage
- Vérification de l'adéquation entre les données d'apprentissages et les données réelles
- Identification des classes rares et stratégies pour les prendre en compte
 - TP : Détection des profils à risque pour le diabète (partie I)
- Réduction de la dimension des données : Analyse en Composantes Principales et Analyse Factorielle
 - TP : Analyse météorologique de l'état du Texas (partie I)
- Clustering, Kmeans
 - TP : Postérisation d'une image

Régression

- Régression, régression linéaire
- Méthode des moindres carrés ordinaire
- Algorithme de Ridge
- Autres algorithmes de régression
 - TP : Analyse météorologique de l'état du Texas (partie II)

Classification par apprentissage supervisé

- Problèmes binaires, Problèmes Multiclasses
- Validation croisée
- SVM
- Random Forest
 - TP : Détection des profils à risque pour le diabète (partie II)

Deep Learning

- Les réseaux de neurones
- Keras et TensorFlow
 - TP : Identification des émotions

Vers des modèles plus complexes

- Apprentissage semi-supervisé et multi-tâches
- Apprentissage par renforcement

ORGANISATION

Formateur

Formation assurée par un expert Data Science

Moyens pédagogiques et techniques

- Accueil des stagiaires dans une salle dédiée à la formation
- Documents supports de formation partagés

- Exposés théoriques
- Étude de cas concrets
- Quiz en ligne
- Mise à disposition en ligne de documents supports à la suite de la formation

Dispositif de suivi de l'exécution de d'évaluation des résultats de la formation

- Feuilles de présence dématérialisées
- Questions orales ou écrites (QCM)
- Mises en situation
- Formulaire d'évaluation de la formation

Délais d'accès

La convocation est envoyée 7 jours avant le début de la formation après réception du bon de commande signé.

Coordonnées de l'équipe pédagogique

- Responsable formation, handicap et votre formateur : Cécile Chardonneau formation@makina-corpus.com
- Suivi facturation : Nathalie Carles Salmon administration@makina-corpus.com