

# PROGRAMME DE FORMATION MACHINE LEARNING

## | Description

Cette formation vous permettra d'acquérir les bases du Machine Learning (apprentissage automatique à partir de données). A la fin des 3 jours, vous saurez créer des modèles prédictifs, les évaluer, et les optimiser. Vous apprendrez également comment les "packager" et les utiliser en production grâce aux APIs.

Tout au long de la formation, vous mettrez en pratique la théorie sur des problématiques de difficulté croissante, en utilisant la plateforme BigML puis la librairie Python scikit-learn (plus bas niveau, mais open source et permettant plus de contrôle). Vous créerez notamment un modèle de prix immobilier, et une API de détection de faux avis pour plateformes hôtelières en ligne.

Si vous êtes déjà à l'aise avec les concepts décrits dans cette formation et que vous souhaitez approfondir certaines notions, notre formation Deep Learning est certainement faite pour vous.

Livre "Bootstrapping Machine Learning" inclus (en anglais)

Les participants à la formation recevront le livre Bootstrapping Machine Learning ainsi que sa version électronique en formats PDF, ePub et Mobi. Ils auront l'occasion, s'ils le souhaitent, de lire le livre avant le début de la formation.

## | Durée

3 jours (21 heures)

## | Objectifs pédagogiques

- Comprendre les possibilités et limites du ML
- Construire des modèles prédictifs à partir de données d'apprentissage, basés sur arbres de décision et ensembles d'arbres
- Interpréter les modèles, analyser leurs comportements, erreurs, performance et impact, pour ensuite les optimiser

- Transformer les variables de type texte en représentations numériques pertinentes pour le ML
- Déployer les modèles en production à l'aide d'APIs

## | Public visé

Cette formation est destinée aux débutants en Machine Learning. Elle sera d'intérêt aux...

- Sociétés souhaitant exploiter la valeur de leurs données et les utiliser dans leurs processus de décision
- Développeurs voulant étendre leur domaine de compétences et apprendre à développer des applications prédictives
- Lead Developers et CTO voulant comprendre les possibilités et enjeux derrière le Machine Learning
- Innovateurs, hackers et startupper voulant tirer profit des possibilités offertes par le Smart Data

Note: nous proposons également une formation Deep Learning destinée aux personnes déjà expérimentées en Machine Learning.

## | Pré-requis

- Bonne compréhension de l'anglais écrit (la formation est en français mais les slides, notebooks, code et autres ressources sont en anglais).
- Connaissances en programmation et de la syntaxe basique du langage Python. Des exemples de code sont fournis tout au long de la formation; les exercices peuvent être réalisés en combinant et en adaptant les briques de code fournies. Consulter Learn Python de Codecademy et Introduction to Python programming de Robert Johansson (en particulier les sections `_Python program files`, `Modules`, `Assignment`, `Fundamental Types`, `Control Flow` et `_Functions_`) pour apprendre ou réviser les bases de Python.
- Utilisation basique d'un tableur (ex: Microsoft Excel).
- Connaissances basiques d'algèbre, de calcul et de statistique appréciables pour une meilleure compréhension de certains modules théoriques.
- Ordinateur portable à apporter.

## | Plan de formation

### Jour 1

Présentation du ML et de ses possibilités

- Concepts de base
- Formalisation des problèmes d'apprentissage supervisé: classification et régression
- Exemples d'utilisation: applications web, mobile, et Data Science en entreprise

#### Création de modèles prédictifs

- Intuitions derrière les techniques d'apprentissage des plus proches voisins (nearest neighbors) et arbres de décision (decision trees)
- Révision des bases de Python; configuration et utilisation de notebooks Jupyter
- Apprentissage de modèles avec la librairie open source scikit-learn et avec l'outil de ML-as-a-Service BigML sur des datasets de classification et de régression
- Visualisation et interprétation

## Jour 2

#### Evaluation des modèles prédictifs

- Critères de performance et procédure d'évaluation
- Mesures agrégées de performance des modèles prédictifs: accuracy, precision, recall, matrices de confusion et de coût (classification), MSE, R-squared, MAPE (regression)
- Mise en pratique avec Python, scikit-learn et BigML sur les datasets précédents
- Procédure d'examen d'erreurs individuelles

#### Optimisation des modèles

- Ajuster la complexité d'un modèle pour éviter le sous-apprentissage / sur-apprentissage
- Comment augmenter l'exactitude des prédictions grâce aux ensembles de modèles; application aux arbres de décision: random forests
- Optimisation de la classification via adaptation du seuil de probabilité, et compromis entre mesures concurrentes
- Procédure de validation et de cross-validation pour benchmarker les modèles
- Procédure de grid-search pour sélectionner les meilleurs paramètres

## Jour 3

#### Utilisation du ML sur du texte — Natural Language Processing

- Pré-traitement des données textuelles avec la librairie NLTK
- Techniques d'extraction de features numériques: sac de mots et n-grams
- Création d'un modèle de détection de faux avis pour plateformes hôtelières en ligne

## Déploiement

- Démonstration de l'utilisation d'APIs REST pour l'utilisation du ML en production
- Déploiement d'une API pour nos propres modèles, avec la librairie Flask
- Utilisation de l'API via curl, Postman, et dans un tableur (remplissage de valeurs manquantes)

## Conclusion

- Récapitulatif des points clé de la formation
- Autres techniques de ML (selon le temps disponible)
  - Initiation aux réseaux de neurones avec le deep learning automatisé sur BigML
  - Apprentissage non supervisé: clustering et détection d'anomalies
  - Prédiction de séries temporelles, via réduction à un problème de régression
  - Systèmes de recommandation, via réduction à un problème de classification
- Ressources pour aller plus loin et suggestions personnalisées

## | Suivi de formation en option

A l'issue de la formation, nos formateurs peuvent aussi intervenir pour vous accompagner dans la mise en application des compétences acquises :

- en répondant à vos questions lors de rendez-vous téléphoniques réguliers
- en étant présents physiquement à l'amorce du projet
- en réalisant un audit de vos pratiques quelques semaines/mois après la formation

*Cette idée vous intéresse ? Faites-le nous savoir à l'acceptation du devis pour que nous trouvions la formule adaptée à votre situation.*

## | Méthodes pédagogiques

On visera une alternance de 50% de travaux pratiques et 50% de cours théoriques. Le support de cours sera fourni au format PDF accompagné d'un lien vers les supports numériques (TP & application).

## | Evaluation des acquis de la formation

Avant le début de la formation, le formateur valide les prérequis et évalue le niveau de départ des participants. L'évaluation des acquis est réalisée tout au long de la formation par des exercices pratiques et des ateliers réalisés par les participants, sous la supervision du formateur.

## I Qualité du/des formateur(s)

Louis Dorard est l'auteur du livre Bootstrapping Machine Learning, du Machine Learning Canvas, General Chair des conférences internationales PAPIs.io, et Adjunct Teaching Fellow à UCL School of Management où il enseigne Predictive Analytics. Dans le cadre de son activité de consultant indépendant, il accompagne grandes entreprises et startups dans l'intégration du ML dans leurs produits. Par le passé, il a "bootstrappé" plusieurs technologies innovantes en tant que Directeur R&D chez Concept Immo Global. Louis est titulaire d'un PhD en Machine Learning de University College London, domaine dans lequel il travaille depuis plus de 10 ans.

Christophe est ingénieur en informatique et en mathématiques appliquées. Après avoir été diplômé de l'Ensimag en 2006, il a été consultant, fondateur de startup, responsable de système d'information, développeur freelance, et a notamment développé des applications de finance personnelle et de Business Intelligence. Au travers de ces expériences, il a pu acquérir une forte culture "data" et a toujours recherché la meilleure technologie pour exploiter la valeur des données.

Christophe met en oeuvre des solutions basées sur le machine learning avec Python, scikit-learn, et diverses plateformes cloud. Formateur passionné et pédagogue, il aime enseigner à des profils variés la puissance de ces techniques quand elles sont mises en oeuvre sur les bons jeux de données. Il est également expérimenté en technologies de bases de données SQL, de traitement de données (Pandas), et web (Django, Selenium).