

# Initiation au Machine Learning

## Durée

3 jours.

## Participants

Ingénieurs/chefs de projet intéressés par les techniques d'apprentissage automatique pour leurs problèmes métiers.

## Prérequis

Bases du langage Python

## Objectifs

Cette formation vous permettra d'acquérir les bases du Machine Learning afin d'identifier les points d'intégrations dans vos applications métier. À la fin de ces trois jours vous serez en mesure de créer 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 en python, utilisant des données issues de problèmes réels et de domaines variés. - Comprendre en quoi consiste l'apprentissage automatique - Identifier les problèmes pour lesquels ce type de méthode saura apporter une réponse pertinente - Connaître les différents types de méthodes et savoir utiliser les principaux algorithmes

## Programme

### Introduction à la modélisation

- Rapide historique du domaine.
- Formalisation d'un problème par le prisme du machine learning.
- Les étapes de construction d'un modèle.
- Les algorithmes supervisés et non-supervisés.

### Évaluation de modèles

- La validation croisée de modèle
- Séparation en base d'apprentissage, de validation et de test.
- Les mesures de performances adaptées à un problème académique ou métier :
  - Précision
  - Rappel
  - Matrice de confusion
  - f-mesures
  - l'aire sous la courbe ROC

### **Les bibliothèques python de calcul scientifique et gestion de corpus**

- Le calcul scientifique avec *numpy*
- Gestion d'un jeu de donnée avec *pandas*
- Affichage de statistiques avec *seaborn*
- Le machine learning avec *Scikit-learn*

### **Les algorithmes non-supervisés**

- De l'intérêt des méthodes non-supervisées
- La réduction de dimension :
  - Les méthodes linéaires liés à l'analyse en composante principale
  - Les méthodes non-linéaires telles que UMAP ou T-SNE
- Le clustering :
  - K-moyennes et dérivées
  - DBSCAN

### **Les algorithmes supervisés**

- Machine a vecteur support (SVM)
- Random Forest
- Réseaux de neurones