

## Formation Machine learning : implémentation en Python

<b>Durée :</b>	5 jours
<b>Public :</b>	Développeurs Python, Data Scientists, Managers de projets
<b>Pré-requis :</b>	Maîtrise des bases du langage Python
<b>Objectifs :</b>	Comprendre et choisir une méthode d'apprentissage automatique - Acquérir les bases du Machine Learning avec Python
<b>Sanction :</b>	Attestation de fin de stage mentionnant le résultat des acquis
<b>Taux de retour à l'emploi:</b>	Aucune donnée disponible
<b>Référence:</b>	INT101603-F
<b>Note de satisfaction des participants:</b>	Pas de données disponibles

### Comprendre l'apprentissage automatique (Machine Learning)

Machine Learning : définition, contextes d'utilisation, phases (apprentissage, mise en production)  
Enjeux et limites  
Approches mathématiques et statistiques  
Application et types de données : graphes, arbres, courbes, vecteurs de caractéristiques  
Qualification de la phase d'apprentissage : classification, régression, renforcement, non supervision  
Panorama d'algorithmes : régressions (linéaire ou logistique), machines à vecteurs, réseaux de neurones, k plus proches voisins (KNN), boosting,...

Facteurs de pertinence et d'efficacité : nombre d'exemples, qualité des attributs, pourcentage des données renseignées ou manquantes, bruit

### Organiser un projet d'apprentissage

Définition du problème  
Acquisition des données  
Analyse et exploration des données  
Préparation et nettoyage des données  
Extraction de caractéristiques  
Choix ou construction du modèle d'apprentissage  
Entraînement, évaluation et optimisation  
Test et vérification de surapprentissage Déploiement

### Découvrir des bibliothèques Python

NumPy : manipulation de matrices et fonctions  
Pandas : lecture et manipulation de données  
Jupyter et ses Notebook : utilisation de cahiers électroniques

### Implémenter des algorithmes d'apprentissage sur des données

Présentation d'une base de données

Régression linéaire (simple ou multiple) : étude, import, création et application d'un modèle, évaluation

Régression logistique : packages, création et application d'un modèle, évaluation et matrice de confusion

K plus proches voisins (KNN) : packages Python, Application et évaluation, amélioration

Support Vector Machine (SVM) : principe, découverte de la base de données, application et évaluation

Analyse en composante principale (PCA) : principe, création du modèle et visualisation des données en 2D

Decision Tree : principe, préparation des données, création du modèle et visualisation de l'arbre de décision

Random Forest : principe, implémentation en Python

### **Evaluer les modèles implémentés**

Ré-échantillonnage

Représentativité des données d'apprentissage Interprétation de la matrice de confusion Sensibilité et spécificité d'un test : Receiver Operating Characteristic (ROC) et Area under the ROC Curve (AUC)