

Plan de cours N° : 1043

Durée : 3 jours (21h)

# Introduction au Machine Learning avec Python

## PARTICIPANTS / PRE-REQUIS

Développeur, chef de projets proche du développement, ingénieur scientifique sachant coder.

Maîtriser l'algorithmique, avoir une appétence pour les mathématiques. La connaissance de Python et des statistiques est un plus.

## OBJECTIFS PEDAGOGIQUES

Appréhender les notions de data science, apprentissage supervisé et non supervisé. Utiliser Python et ses bibliothèques pour manipuler des données. Créer et gérer des DataLake, DataMart et DataWarehouse. Comprendre et utiliser les différents types de régression, notamment la régression linéaire. Maîtriser Matplotlib pour afficher des données sous forme de graphiques 2D. Manipuler des DataFrames et importer des données CSV et SQL. Créer, entraîner et évaluer des modèles de régression et de classification. Appliquer l'algorithme des k plus proches voisins (k-NN) et d'autres modèles comme SVM et Random Forest. Appréhender les concepts de perceptron, réseaux neuronaux MLP et descente du gradient. Utiliser des bibliothèques pour créer et optimiser des réseaux de neurones, y compris les réseaux convolutifs et le transfert learning.

## MOYENS PEDAGOGIQUES

Tour de table au début de chaque formation pour définir les objectifs de chaque participant,

Alternance entre apports théoriques (en moyenne 30%) et exercices pratiques (en moyenne 70%),

Utilisation de cas concrets issus de l'expérience professionnelle de nos formateurs,

Remise d'un support de cours,

Assistance post-formation d'une durée de 1 an sur le contenu de la formation via notre adresse mail dédiée [formateurs@atp-formation.com](mailto:formateurs@atp-formation.com)

## MOYENS PERMETTANT LE SUIVI DE L'EXECUTION ET DES RESULTATS

Positionnement préalable oral ou écrit,

Evaluation des acquis tout au long de la formation par des exercices de synthèse,

Attestation de stage remise à chaque apprenant, avec son niveau d'acquisition pour chaque objectif pédagogique,

Feuille de présence signée par demi-journée,

Questionnaire de satisfaction pour évaluer la qualité de l'enseignement,

En option : passage certification possible selon les thématiques.

## MOYENS TECHNIQUES EN PRESENTIEL

Accueil des stagiaires dans une salle dédiée à la formation, équipée d'ordinateurs récents et performants, d'un vidéo projecteur et d'un tableau blanc.

## MOYENS TECHNIQUES DES CLASSES A DISTANCE

Grâce à un logiciel comme Teams, suivez une formation en temps réel et entièrement à distance. Lors de la classe en ligne, les apprenants interagissent et communiquent entre eux et avec le formateur.

Nous vous conseillons très fortement l'utilisation de votre webcam et de disposer d'un double écran.

Pour toute question avant et pendant le parcours, une assistance technique et pédagogique est à disposition par téléphone au 04.76.41.14.20.

## ORGANISATION

Les cours ont lieu de 9h00-12h30 13h30-17h00 (adaptable à la demande).

## PROFIL FORMATEUR

Nous recrutons méticuleusement nos formateurs selon 3 critères: expertise, pédagogie et agilité.

## ACCESSIBILITE

Les personnes atteintes de handicap souhaitant suivre nos formations sont invitées à nous contacter directement, afin d'étudier ensemble les possibilités d'organisation.

## MISE A JOUR

15/07/2024

### Siège social :

31 avenue du Granier  
38240 MEYLAN

### Agences :

170 rue de Chatagnon  
38430 Moirans

### Le Thélème

1501/1503 route des Dolines  
06560 Valbonne

Plan de cours N° : 1043

Durée : 3 jours (21h)

# Introduction au Machine Learning avec Python

## INTRODUCTION AUX DATA SCIENCES

- Qu'est-ce que la data science ?
- Qu'est-ce que Python ?
- Qu'est-ce que le Machine Learning ?
- Apprentissage supervisé vs non supervisé
- Les statistiques
- La randomisation
- La loi normale

## RAPPELS DE PYTHON POUR LES DATA SCIENCES

- Rappel
- PIP

## INTRODUCTION AUX DATALAKE, DATAMART ET DATA WHAREHOUSE

- Qu'est-ce qu'un DataLake ?
- Les différents types de DataLake
- Le Big Data
- Qu'est-ce qu'un DataWarehouse ?
- Qu'est-ce qu'un DataMart ?
- Mise en place d'un DataMart

## MACHINE LEARNING

- Filtrage
- Mise en place d'une machine learning supervisée
- Qu'est-ce qu'un modèle et un dataset
- Qu'est-ce qu'une régression
- Les différents types de régression
- La régression linéaire
- Gestion du risque et des erreurs
- Quarter d'Ascombe
- Trouver le bon modèle
- La classification
- Loi normale, variance et écart type
- Apprentissage
- Mesure de la performance
- No Fee Lunch
- Les pièges du machine learning : Overfitting, Underfitting

## MATPLOTLIB

- Utilisation de la bibliothèque scientifique de graphe Matplotlib
- Affichage des données dans un graphique 2D
- Affichage des sous-graphes
- Affichage des polynômes et de sinusoidales

## PANDAS

- Le Data Mining
- L'analyse des données avec Pandas
- Les DataFrames
- La théorie ensembliste avec Pandas
- L'importation des données CSV
- L'importation des données SQL
- Pandas et SKLearn

## SCIKITLEARN

- Le machine Learning avec SKLearn
- La régression linéaire
- La création du modèle
- L'échantillonnage
- La randomisation
- L'apprentissage avec fit
- La prédiction du modèle
- Les metrics
- Choix du modèle
- PreProcessing et Pipeline
- Régressions non polynomiales

## NEAREST NEIGHBORS

- Algorithme des k plus proches voisins (k-NN)
- Modèle de classification
- K-NN avec SciKitLearn
- Choix du meilleur k
- Sérialisation du modèle
- Variance vs Erreurs
- Autres modèles : SVN, Random Forest
- La clusterisation avec k-Means

### **Siège social :**

31 avenue du Granier  
38240 MEYLAN

### **Agences :**

170 rue de Chatagnon  
38430 Moirans

### **Le Thélème**

1501/1503 route des Dolines  
06560 Valbonne

Plan de cours N° : 1043

Durée : 3 jours (21h)

# Introduction au Machine Learning avec Python

## SCALABLE VECTOR MACHINE

- Le modèle SVM
- SVM et les petits datasets
- Choix du kernel

## RANDOM FOREST

- Le principe de RF
- Principe de la boîte blanche
- Reverse engineering RF
- Les corrélations automatiques

## LES RESEAUX NEURONAUX

- Le perceptron
- Les réseaux neuronaux MLP
- La descente du gradient

## TENSORFLOW ET KERAS

- Les réseaux Denses
- Keras
- Dimensionnement du réseau
- Les réseaux convolutifs
- Le transfert learning
- Le réapprentissage
- L'apprentissage par renforcement
- Les modèles courants
  - VGG
  - RestNet
  - Xception
  - LSTM
  - BERT
  - MobileNet
- L'optimisation des réseaux

### **Siège social :**

31 avenue du Granier  
38240 MEYLAN

### **Agences :**

170 rue de Chatagnon  
38430 Moirans

### **Le Thélème**

1501/1503 route des Dolines  
06560 Valbonne