

Mis à jour le 05/02/2025

S'inscrire

# Formation Python pour les Scientifiques avec Scikit-Learn

2 jours (14 heures)

## Présentation

[Scikit-learn](#) (ou Sklearn) est une bibliothèque libre Python dédiée à l'apprentissage automatique. Elle est développée par de nombreux contributeurs notamment par des instituts français d'enseignement supérieur et de recherche comme Inria et Télécom ParisTech.

Elle comprend notamment des fonctions pour estimer des forêts aléatoires, des régressions logistiques, des algorithmes de classification, et les machines à vecteurs de support. Elle est conçue pour s'harmoniser avec d'autres bibliothèques libres Python, notamment NumPy et SciPy.

Cette formation Python pour les scientifiques vous permettra d'acquérir les connaissances nécessaires pour interagir avec Sklearn. Elle abordera sa syntaxe, les outils ainsi que les bonnes pratiques de développement, afin de bénéficier de tous les atouts de cette bibliothèque. Nous présenterons les algorithmes, les méthodes supervisées et non supervisées, les fonctionnalités de classification, régression et le clustering.

Comme toutes nos formations, celle-ci vous présentera la dernière version stable en date et ses nouveautés ([Scikit-learn 1.6](#) sortie en Juin 2023 à la date de l'article, et [Python 3.13](#)).

## Objectifs

- Mettre en œuvre des algorithmes d'apprentissage machine en Python avec Scikit-Learn
- Pouvoir choisir une stratégie de montée en charge pour Scikit-Learn
- Utiliser Scikit-Learn en conjonction avec d'autres toolkit de l'univers Python (par ex. Skimage ou OpenCV)
- Être en mesure de citer les différents composants de Scikit-Learn, leur usage, trouver rapidement leur documentation

## Public visé

- Data analysts
- Data Miners
- Statisticiens
- Scientifiques

## Pré-requis

- Savoir programmer en Python 2 ou 3
- De bonnes connaissances en traitement et en manipulation de données
- De bonnes connaissances en algèbre linéaire

## Programme de la formation Python Scientifique avec Scikit.Learn

### Introduction

- Qu'est-ce que le Machine Learning ?
- Apprentissage Machine Learning ?
- Installation scikit-learn
- Clean Data
- Sur-ajustement et la généralisation
- Validation croisée
- Bias-Variance Trade-Off
- Vérification et mise à jour de l'installation de scikit-learn
- Solutions et Avantages : L'état des arts
  - Tensorflow et Keras
  - Scikit?learn
  - Deeplearning for Java
  - pytorch et fastai

### Concepts de bases de SCIKIT

- Algorithmes, vocabulaire
- Panorama de scikit-learn
- Démarche d'apprentissage
- Périmètre du toolkit, domaine d'application
- Structure
- Représentation des données et principes de l'API
- La documentation

### MÉTHODES DE MACHINE LEARNING

- Les méthodes supervisées
  - KNN
  - SVM
  - Réseaux de neurones
  - Réseau bayésien naif
  - Régression pénalisée
  - Boosting
  - Random forest
- Les méthodes non supervisées
  - Clustering
  - Règles d'associations et de séquences
  - Traitement du texte (text mining)

## Catégories de problèmes et mise en oeuvre

- Classification
  - Définition des arbres de décision
  - Arbres de décision
  - Modèles d'ensemble
  - Classificateur de Random Forest
- Régression
  - Arbres de décision
  - Modèles de régression
  - Définition d'un hyperplan
  - Anatomie d'un estimateur
  - Régression linéaire simple
  - La fonction COST
  - Qu'est-ce que R-Squared ?
  - Évaluation du modèle
  - Régularisation
  - Régression linéaire à plusieurs variables
  - Régression linéaire appliquée
  - Visualisation de notre DataSet
- Clustering
  - Animation du Cluster K-Means
  - L'algorithme K-Means
  - Application du Clustering K-Means
- Autres algorithmes disponibles

## Les modèles

- Validation des modèles
- Préparation des modèles
  - Cas des images
  - Cas des textes
  - Réduction des dimensions
  - Autres préparations

## Optimisation

- Comment régler son algo de ML ?
- HPO = Hyper Parameter Optimization
- AutoML
- Présentation d'un kaggle master
  - Utilisation d'optimisation bayésienne

## Utilisation des paramètres de gestion SCIKIT

- Choisir le bon estimateur et les bons paramètres
- Interopérabilité avec écosystème python scientifique
- Performance et montées en charge
  - Stratégie
  - Outils

## Sociétés concernées

Cette formation s'adresse à la fois aux particuliers ainsi qu'aux entreprises, petites ou grandes, souhaitant former ses équipes à une nouvelle technologie informatique avancée ou bien à acquérir des connaissances métiers spécifiques ou des méthodes modernes.

## Positionnement à l'entrée en formation

Le positionnement à l'entrée en formation respecte les critères qualité Qualiopi. Dès son inscription définitive, l'apprenant reçoit un questionnaire d'auto-évaluation nous permettant d'apprécier son niveau estimé sur différents types de technologies, ses attentes et objectifs personnels quant à la formation à venir, dans les limites imposées par le format sélectionné. Ce questionnaire nous permet également d'anticiper certaines difficultés de connexion ou de sécurité interne en entreprise (intraentreprise ou classe virtuelle) qui pourraient être problématiques pour le suivi et le bon déroulement de la session de formation.

## Méthodes pédagogiques

Stage Pratique : 60% Pratique, 40% Théorie. Support de la formation distribué au format numérique à tous les participants.

## Organisation

Le cours alterne les apports théoriques du formateur soutenus par des exemples et des séances de réflexions, et de travail en groupe.

## Validation

À la fin de la session, un questionnaire à choix multiples permet de vérifier l'acquisition correcte des compétences.

## Sanction

---

Une attestation sera remise à chaque stagiaire qui aura suivi la totalité de la formation.

[Page Web du Programme de Formation](#) - Annexe 1 - Fiche formation

Organisme de formation enregistré sous le numéro 11 75 54743 75. Cet enregistrement ne vaut pas agrément de l'État.

© Ambient IT 2015-2025. Tous droits réservés. Paris, France - Suisse - Belgique - Luxembourg