

PYTHON

L'apprentissage Machine en Python (SciKit-Learn)

2 jours (14h00) | PYT005 | Num form : form-224 | Perfectionnement / Avancé

INFORMATIQUE / DÉVELOPPEMENT / PYTHON

À l'issue de ce stage vous serez capable de :

- Être capable de mettre en oeuvre SciKit-learn pour de l'apprentissage machine et l'analyse de données
- Comprendre comment collecter les données et les préparer pour un traitement efficient
- Savoir modéliser les données et optimiser les modèles
- Être en mesure d'industrialiser et de déployer une application de prédiction

Niveau requis :

- Connaissance de Python et d'une bibliothèque de calcul telle que Numpy ou Pandas

Public concerné :

- Chefs de projet et data-scientists souhaitant comprendre le fonctionnement de SciKit-learn

Programme :

L'ÉCOSYSTÈME SCIKIT LEARN

- Origine
- Missions et évolutions
- Architecture
- Modules
- Atelier pratique : Installation de l'écosystème (PC, MAC, LINUX)

BONNES PRATIQUES SCIKIT LEARN : LA RÉCUPÉRATION DE DONNÉES

- L'API dédiée aux jeux de données
- Problématique des grands jeux de données
- Les jeux de données internes au framework
- La génération de données artificielles
- L'accès aux données ouvertes
- Méthodologies de chargement de données externes
- Atelier pratique : Collecte de données sur un cas concret

BONNES PRATIQUES SCIKIT LEARN : LE PRÉTRAITEMENT

- Cadre et rôle du prétraitement
- Méthodes de mise à l'échelle des données
- Normalisation des données
- Traitement des données catégorielles
- Traitement des données manquantes
- Atelier pratique : Prétraitement des données dans un cas concret

BONNES PRATIQUES SCIKIT LEARN : L'INGÉNIERIE DES VARIABLES PRÉDICTIVES

- Cadre et rôle de l'ingénierie des variables prédictives
- Transformations non linéaires
- Algorithmes de réduction dimensionnelle de données
- L'extraction automatique de traits (textes, images)
- Combinaison et transformations ad-hoc des données
- Atelier pratique : Ingénierie des variables prédictives sur un cas concret

BONNES PRATIQUES SCIKIT LEARN : LA MODÉLISATION

- Cartographie des algorithmes de l'écosystème
- Tour de l'apprentissage supervisé avec SciKit Learn
- Tour de l'apprentissage non supervisé avec SciKit Learn
- En pratique, mode de sélection d'algorithmes pertinents
- Atelier pratique : L'apprentissage supervisé et non supervisé sur cas concrets

BONNES PRATIQUES SCIKIT LEARN : SÉLECTION ET OPTIMISATION DES MODÈLES

- Présentation et usage des outils de combinaison et de chaînage de processus
- L'évaluation de performance par validation-croisée

- Le réglage des hyper-paramètres d'un modèle
- Les APIs d'évaluation des modèles de prédiction
- Les différentes courbes d'évaluations de modèles de prédictions
- Atelier pratique : Sélection de modèle, et optimisation des réglages

BONNES PRATIQUES SCIKIT LEARN : INDUSTRIALISATION ET DÉPLOIEMENT

- Persistance des modèles
- Scaling des applications de prédictions avec SciKit Learn
- Problématiques de latence et de débit à l'exécution
- La parallélisation
- La gestion de l'environnement Python associé
- Atelier pratique : Industrialisation d'une application de prédiction