

Machine Learning avec scikit-learn

Référence : PYDS011

Durée : 2 jours (14 heures)

Certification : Aucune

CONNAISSANCES PRÉALABLE

- Connaissance de python et d'une bibliothèque de calcul telle que numpy ou pandas

PROFIL DES STAGIAIRES

- Chefs de projet, data-scientists, souhaitant comprendre le fonctionnement de scikit-learn

OBJECTIFS

- Mettre en oeuvre scikit-learn pour de l'apprentissage machine et l'analyse de données

CERTIFICATION PRÉPARÉE

- Aucune

MÉTHODES PÉDAGOGIQUES

- 6 à 12 personnes maximum par cours, 1 poste de travail par stagiaire
- Remise d'une documentation pédagogique papier ou numérique pendant le stage
- La formation est constituée d'apports théoriques, d'exercices pratiques et de réflexions

FORMATEUR

- Consultant-Formateur expert Bigdata

MÉTHODES D'ÉVALUATION DES ACQUIS

- Auto-évaluation des acquis par le stagiaire via un questionnaire
- Attestation des compétences acquises envoyée au stagiaire
- Attestation de fin de stage adressée avec la facture

ACCESSIBILITÉ DE LA FORMATION

- EduGroupe met en place un ensemble de dispositifs pour accueillir, accompagner et adapter ses formations aux personnes en situation de handicap (PSH).
- Notre référent(e) handicap se tient à votre disposition au 01.71.19.70.30 ou par mail à referent.handicap@edugroupe.com pour tout besoin d'aménagement, afin de vous offrir la meilleure expérience possible.

CONTENU DU COURS

1. Présentation

- Historique
- Fonctionnalités
- Lien avec Numpy et Scipy

2. Manipulation de données

- Chargement de données
- Pré-traitement de données : standardisation, transformations non linéaires, discrétisation
- Génération de données

3. Analyse des données et classification

- Modèles : linéaires, quadratiques, descente de gradient
- Algorithmes, choix d'un estimateur
- Classification : k-voisins, régression logistique, classification naïve bayésienne, arbres de décision, forêts aléatoires, SVC
- Régression : régression linéaire, lasso, SGDr, SVR
- Détection de groupes : k-moyennes, Spectral Clustering/GMM
- Analyse globale : Randomized PCA, kernel approximation
- Exemple de Travaux Pratiques : classification automatique d'un jeu de données à partir d'une régression logistique
- Création de jeux d'essai, entraînement et construction de modèles
- Prévisions à partir de données réelles. Mesure de l'efficacité des algorithmes. Courbes ROC
- Parallélisation des algorithmes. Choix automatique
- Exemple de Travaux Pratiques : mise en évidence des erreurs d'apprentissage en fonction des hyper-paramètres
- Recherche de clusters : modélisations, algorithmes, et méthodes d'évaluation
- Réseaux de neurones

4. Modèles d'apprentissage

- Chargement et enregistrement
- Génération de modèles
- Estimation de la performance d'un modèle
- Mesures de performance
- Modification des hyper-paramètres
- Application pratique avec les courbes d'évaluations