



Machine Learning avec scikit-learn

DS011

Durée: 2 jours

1 610 €

12 au 13 février

8 au 9 avril
17 au 18 juin

16 au 17 septembre
14 au 15 novembre

Public :

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

Connaissances préalables nécessaires :

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

Programme :

Présentation

Historique
Fonctionnalités
Lien avec Numpy et Scipy

Manipulation de données

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

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

Atelier : 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.



Phirio

Atelier : Mise en évidence des erreurs d'apprentissage en fonction des hyper-paramètres

Recherche de clusters: modélisations, algorithmes, et méthodes d'évaluation
Réseaux de neurones

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