



Machine learning avec Python Scikit Learn

Référence et durée

PYTHML

2 jours, 14 heures

Objectifs

Avoir une vision transverse des principaux algorithmes et méthodes d'analyse prédictive supervisée. Savoir mettre en œuvre les algorithmes dans Scikit-Learn sur un problème de classification ou de régression

Public et prérequis

Cette formation s'adresse à un public d'analystes, data miners, chargés d'études statistiques, directeurs d'études. Des connaissances statistiques et informatiques sont préférables. Avoir suivi le cours [Traitement et analyse de données dans Python](#) ou avoir connaissance de l'environnement Python.

Moyens pédagogiques

Alternance de présentations théoriques et de démonstrations avec des exercices pratiques en programmation Python réalisés par les participants. Les exercices permettent de valider les acquis.

Programme détaillé

- Introduction
 - Algorithmes, vocabulaire
 - Panorama de scikit-learn
 - Démarche d'apprentissage
- Démarche d'apprentissage détaillée
 - Feature extraction (décomposition de matrices, n-grams, hashing, tf-idf)
 - Echantillonnage, validation, cross-validation, grid search, feature selection, metriques de validation (mape, AUC, ...), biais-variance, erreur d'apprentissage, erreur réelle
- 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)
- Mise en pratique d'algorithmes non supervisés
 - Analyse de données textuelles : feature extraction, analyse latente
- Mise en pratique d'algorithmes supervisés
 - Exercice de mise en œuvre d'un modèle (apprentissage supervisé d'une cible binaire ou continue) à partir d'un grand nombre de variables
 - Métriques d'Evaluation et comparaison de modèles