

Exploitation des données massives en finance

Infos pratiques

- > ECTS : 1.5
- > Nombre d'heures : 16.0
- > Langue(s) d'enseignement : Français
- > Période de l'année : Enseignement dixième semestre
- > Méthodes d'enseignement : En présence
- > Forme d'enseignement : Travaux dirigés
- > Ouvert aux étudiants en échange : Oui
- > Composante : Sciences économiques, gestion, mathématiques et informatique
- > En savoir plus : <https://sites.google.com/view/mastergda>

Présentation

- Définition du concept de données massives. Concept des 3Vs.
- Présentation de sources de données boursières et financières gratuites ou commerciales
- Récupération de données disséminées sur le WEB à l'aide d'outils de Scraping
- De la donnée brute à la donnée exploitable : préparation des données comme input de modèles
- Introduction concept d'apprentissage statistique, classification, régression, panoplie de modèles existants
- Régularisation des données : overfitting, underfitting, hyperparamètre et gestion de la variance
- Arbres de décisions, forêts aléatoires
- Introduction aux Réseaux de Neurones Artificiels (ANN) et leurs dérivés (RNN, ...)

Objectifs

L'objectif du présent cours-TD est double. A l'aide de Python, nous explorerons d'abord les problématiques liées à la gestion de données massives (Big Data). Une fois ces données retraitées convenablement, nous

présenterons quelques méthodes de Machine-Learning utiles pour la régression et classification.

Le cours prendra la forme d'une conférence interactive (comme une PyConf).

Évaluation

Évaluation écrite (partiel traditionnel)

Pré-requis nécessaires

Compétence intermédiaire ou avancée en programmation Python et R

Compétences visées

Python : utilisation avancée des packages Pandas & Scikit Learn

Finance : connaissance de l'éventail des bases de données disponibles

Scraping : Extraction de données disséminées sur le WEB

BIG data : De la donnée brutes à la donnée exploitable

ML : Arbre de décision, Forêt Aléatoire, ANN, RNN

Bibliographie

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Richert, W. (2013). *Building machine learning systems with Python*. Packt Publishing Ltd

Harrington, P. (2012). *Machine learning in action*. Manning Publications Co..

Ressources pédagogiques

Classe interactive