

Plan de cours 2021-2022
Nom de l'unité de cours : Data Mining
Nom du professeur : Alexandros Kalousis
Semestre d'automne

Objectifs du cours

Objectifs principaux:

- comprendre les principales tâches de data mining et comment ils peuvent être abordés par les différents algorithmes de data mining,
- comprendre le fonctionnement interne des algorithmes de data mining et sous quelles conditions ils devraient être utilisés,
- apprendre à construire des workflows complexes de data mining et des processus et comment évaluer les entre eux

Objectifs opérationnels :

- maîtriser le langage d'analyse R
- être en mesure de aborde des problèmes réels et complexes du data mining

Renseignements pratiques

- Les cours, théorie et pratique, ont lieu le Lundi (17:15-19:45), B3.12
Responsable: Alexandros.Kalousis@hesge.ch, Bureau F124, 1re étage
Assistants :
 - Frantzeska.Lavda@hesge.ch, Bureau F107, 1re étage
 - Yoann.Boget@hesge.ch, Bureau F107, 1^{re} etage

Mode d'évaluation

- Première tp obligatoire, mais non note.
- Trois TPs à remettre en séance de TP (travail à deux), 30% de la note finale.
- Examen final:
 - un projet à remettre avant la fin du semestre d'automne suivi par la présentation du projet (travail à deux). Le projet sera la synthèse de votre travaille sur le trois TP plus l'équivalent des deux TPs.
- La note finale est la moyenne pondérée des
 - notes des exercices (30%)
 - et du devoir du semestre d'automne ou d'examen écrit (70%).

Plan du cours (les sujets peuvent être abordés dans un ordre différent)

Semestre d'automne

1. Introduction au data mining
2. L'apprentissage supervisé (classification, régression)
 - Le partitionnement récursif : Arbres et règles de décision: C4.5, CART
 - L'apprentissage par estimation de densité: méthodes bayésiennes, voisins les plus proches, kNN,
 - Approches numériques linéaires: discriminants linéaires, SVM
 - Approches numériques non linéaires: méthodes à base de noyaux (SVM)
3. Techniques d'évaluation et d'expérimentation
 - Mesures de performances et stratégies d'expérimentation
 - Comparaison et sélection de modèles
4. Réduction des données
 - Réduction de la dimensionnalité: sélection/transformation de variables

Méthodes pédagogiques

Le contenu de l'unité de cours est offert chaque semaine pendant trois heures d'enseignement (théorie

et pratique). Le cours demande de la part des étudiants une forte implication quant à la réalisation des exercices. Il convient donc que chacun s'investisse de manière adéquate et que les travaux soient réalisés régulièrement. L'ensemble des supports et des énoncés mis à disposition est disponible sur le Tableau de bord du cours sous **Cyberlearn**: <http://cyberlearn.hes-so.ch>.

Références

- Ian H. Witten, Eibe Frank, Mark A. Hall. Data Mining: Practical Machine Learning Tools and Techniques (Third Edition). Morgan Kaufmann, 2011.
- Bishop, C. M. Pattern Recognition and Machine Learning. Springer, 2006.
- Hastie, T., Tibshirani, R, Friedman, J. Elements of Statistical Learning: Data Mining, Inference and Prediction. Springer, 2001.
- Mitchell, T. Machine Learning. McGraw-Hill, 1997.