

## Plan de cours 2018-2019

**Nom du module : 625-2 - Valorisation des systèmes d'information**

**Nom de l'unité de cours : Datamining**

**Nom du professeur : Magda Gregorová**

**Semestre automne**

### Objectifs du cours (en conformité avec le descriptif de module)

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 les bases de langage Python3
- être en mesure de aborde des problèmes réels et complexes du data mining

### Plan de cours (Contenu détaillé du cours)

Concepts de base de l'analyse de données

- types de données, prétraitement des données, exploration des données, visualisation des données
- concepts de probabilités élémentaires
- mesures de similarité, distances, dépendance

Condept de base d'apprentissage supervisé

- modélisation, évaluation et sélection
- mesures de performance et tests

Algorithmes de classification sélectionnés:

- sélection de fonctionnalités pour la classification
- arbres de décision
- voisin le plus proche
- Bayes naïfs
- discriminantes linéaires

Python pour l'analyse de données

### 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>.

### Mode d'évaluation (en conformité avec le descriptif de module)

- 30 % trois TP (chaque 10 %) - groupes de 3 personnes, analyse complète (codage et commentaires)
- 70 % examen final - mêmes groupes, même type d'analyse que dans les TP mais sur des données différentes, préparation 1 heure + 30 minutes orales
- (+/- 5 %) participation dans le cours (exercice, foire aux questions, etc.)

Planification : (sous réserve de modification)

- TP1 à remettre 28/10/2018
- TP2 à remettre 25/11/2018
- TP3 à remettre 6/1/2019

### Bibliographie

- Zaki, M.J. & Wagner, M.J, Data Mining and Analysis: Fundamental Concepts and Algorithms, 2012 Cambridge University Press
- Han, J., Kamber, M. & Pei, J., Data Mining: Concepts and Techniques, 2011 Morgan Kaufmann
- Tan, P.-N., Steinbach, M. & Vipin K., Introduction to data mining, 2005 Wesley Longman Publishing Co
- Witten, I.H., Frank, E. & Hall, M.A, Data Mining: Practical Machine Learning Tools and Techniques
- McKinney, W., Python for Data Analysis, 2018 O'Reilly
- Guido, S., Mueller, A., Introduction to Machine Learning with Python, 2016 O'Reilly