

# Informations

## Généralités

L'option mineure de Data mining se déroule le jeudi soir de 18h15 (ou 17h15) jusqu'à 21h30. Tout retard non justifié sera sanctionné.

Cette option est divisée en deux parties

- La première partie sera dédiée plus spécifiquement à la théorie, ainsi qu'à la présentation de cas réels. Elle comporte 4 séances.
- La deuxième partie sera dédiée à l'apprentissage de Clementine, un logiciel professionnel de data mining.

Le site <http://campus.hesge.ch/varones/> onglet "Data Mining" est le site dédié à ce cours.

## Travaux pratiques et examen

Deux travaux pratiques notés devront être faits.

1. Une présentation d'une méthode de data mining sera demandée. Ce travail se fera en binôme, avec une présentation orale et la remise d'un rapport écrit ( max 5 pages). La présentation orale aura lieu le 5 mars et le 2 avril, durera 20 minutes, et sera suivie de 10 minutes pour les réponses aux questions. Le rapport écrit devra être rendu au plus tard le **16** avril. Les sujets disponibles sont donnés ci-dessous :
2. Un travail pratique individuel sera évalué sur la base d'un rapport (10 à 15 pages). Il consistera en un processus complet de Data Mining, utilisant le logiciel Clementine. Le rapport devra être rendu pour le ~~28 mai~~ 4 juin.
3. Un examen aura lieu durant la semaine 16.

La note finale de cette option se composera de

- 20% Présentation orale d'une méthode de data mining
- 20% Rapport écrit sur une méthode de data mining
- 40% Travail pratique
- 20% Examen.

## Liste de sujets

La formation des binômes, ainsi que le sujet choisi est annoncée par les étudiants au plus tard le jeudi 26 février. En cas de choix identique par plusieurs binôme, l'attribution des sujets se fera selon la règle "FIFO" (first in, first out), i.e. le premier binôme qui s'annonce pour un sujet se verra attribué le sujet en question.

Le chiffre dans le carré indique le nombre de personnes ayant choisi le sujet. Un chiffre rouge indique que le sujet n'est plus disponible.

1. 0 Réduction de données : Analyse en Composantes Principales
2. 0 Réduction de données : Analyse Factorielle des Correspondances
3. 0 Réduction de données : Analyse des Correspondances Multiples
4. 2 Classification : Arbres de Décision
5. 2 Classification : k-means,  $k$  plus proches voisins
6. 0 Classification : Support Vector Machine
7. 0 Classification : Regression Logistique
8. 2 Analyse discriminante
9. 2 Classification : Ascendante hiérarchique
10. 2 Classification : Naive Bayes Model
11. 2 Segmentation : Règles d'association (Apriori, ...)
12. 0 Segmentation : Réseaux de kohonen
13. 2 Prédiction : Réseaux neuronaux
14. 0 Prédiction : Régression multiple
15. 2 Scoring : Ranking de pages web avec PageRank

Un choix de sujet hors liste reste toutefois possible, pour autant que le sujet soit accepté par le responsable de l'option.

## Ordre de présentation

5.03.09 5, 4, 9, 15

2.04.09 11, 8, 10, 13