

Institut de statistique



STAT2550 Data Mining

[15h+15h exercices] 5 crédits

Cette activité se déroule pendant le 2ème semestre

Enseignant(s): Libei Chen
Langue d'enseignement : français
Niveau : Deuxième cycle

Objectifs (en termes de compétences)

Au terme du cours l'étudiant aura acquis des connaissances de base sur la méthodologie et les outils du data mining pour le traitement de grandes bases de données d'entreprises. Il aura conscience des spécificités et de l'apport du data mining par rapport à la statistique classique et sera capable de traiter des études de cas avec un logiciel dédié.

Objet de l'activité (principaux thèmes à aborder)

- Domaines d'application du data mining.
- Etapes types de la méthodologie du data mining.
- Echantillonnage et partitionnement de la base de données.
- Prétraitement et validation des données.
- Exploration, réduction et transformation des variables.
- Outils de modélisation et de classification du data mining (présentation générale).
- Arbres de décision.
- Réseaux de neurones.
- Outils de validation du modèle.
- Etudes de cas.

Résumé : Contenu et Méthodes

- Introduction au data mining
 - o Data et systèmes du data mining
 - o Domaines d'application du data mining
 - o Processus et méthodologie du data mining
 - o Data mining dans le domaine du Customer Relationship Management (CRM)
 - o Statistique classique et data mining
- Préparation des données pour data mining
 - o Principales étapes dans la préparation des données
 - o Spécification des données pour le data mining
 - o Extraction et agrégation des données
 - o Audit et exploration des données
 - o Pré-traitement des données
- Modélisation prédictive
 - o Arbres de décision
 - o Réseaux de neurones
 - o Validation et choix de modèles
- Modélisation descriptive
 - o Clustering
 - o K-means
 - o Kohonen Self-Organising Map
- Etudes de cas

Autres informations (Pré-requis, Evaluation, Support, ...)

Ouvrages de référence

1. Berry M. and G. Linoff (2000), "Mastering Data Mining, The Art and Science of Customer Relationship Management", John Wiley.
2. Bishop, C.M. (1995), Neural Networks for Pattern Recognition, Oxford.
3. Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. (1984), "Classification and Regression Trees", Wadsworth, Inc., Belmont, California.
4. Han J. and M. Kamber (2000), "Data Mining: Concepts and Techniques", Morgan Kaufmann,.
5. Hastie Tr., R. Tibshirani and J. Friedman (2001), "The Elements of Statistical Learning -Data Mining, Inference and Prediction", Springer.
6. Haykin S., "Neural Networks: A comprehensive Foundation", Prentice Hall, 1999
7. Kohonen T. (1995), "Self-Organizing Maps", Springer Series in Information Sciences, Oxford University Press.
8. Piatetsky-Shapiro G. and W. J. Frawley (1991), "Knowledge Discovery in Databases", AAAI/MIT Press.
9. Piatetsky-Shapiro G., U. Fayyad, and P. Smith (1996). "From data mining to knowledge discovery: An overview", In U.M. Fayyad, et al. (eds.), Advances in Knowledge Discovery and Data Mining, 1-35. AAAI/MIT Press,.
10. Pyle D. (2000), "Data Preparation for Data Mining", Morgan Kaufman.
11. Richard O. Dula, Pete E. Hart and David G. Stork (2000), "Pattern Classification", John Wiley, Second edition.
12. Van Hulle M. (2000), "Faithful Representations and Topographic Maps: From Distortion- to Information-Based Self-Organization", John Willey

Autres crédits de l'activité dans les programmes

ECGE3DS/MK	Diplôme d'études spécialisées en économie et gestion (Master in business administration) (marketing)	(5 crédits)	Obligatoire
ECGE3DS/SC	Diplôme d'études spécialisées en économie et gestion (Master in business administration) (Supply Chain Management)	(5 crédits)	Obligatoire
STAT21MS/DM	Première année du master en statistique, orientation générale, à finalité spécialisée (data management et data mining)	(5 crédits)	Obligatoire
STAT21MS/EA	Première année du master en statistique, orientation générale, à finalité spécialisée (économie et assurance)	(5 crédits)	
STAT21MS/MS	Première année du master en statistique, orientation générale, à finalité spécialisée (marketing et sondage)	(5 crédits)	
STAT22MS/DM	Deuxième année du master en statistique, orientation générale, à finalité spécialisée (data management et data mining)	(5 crédits)	Obligatoire
STAT22MS/EA	Deuxième année du master en statistique, orientation générale, à finalité spécialisée (économie et assurance)	(5 crédits)	
STAT22MS/MS	Deuxième année du master en statistique, orientation générale, à finalité spécialisée (marketing et sondage)	(5 crédits)	
STAT3DA/E	diplôme d'études approfondies en statistique (statistique et économétrie)	(5 crédits)	
STAT3DA/P	diplôme d'études approfondies en statistique (pratique de la statistique)	(5 crédits)	