UCL Université catholique de Louvain

LSTAT2350

2015-2016

Data Mining

5.0 crédits	15.0 h + 15.0 h	2q

Enseignants:	Chen Libei ;
Langue d'enseignement:	Français
Lieu du cours	Louvain-la-Neuve
Préalables :	*Le(s) prérequis de cette Unité d'enseignement (UE) sont précisés à la fin de cette fiche, en regard des programmes/formations qui proposent cette UE.*
Thèmes abordés :	- 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.
Acquis d'apprentissage	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 dédicacé. *La contribution de cette UE au développement et à la maîtrise des compétences et acquis du (des) programme(s) est accessible à la fin de cette fiche, dans la partie « Programmes/formations proposant cette unité d'enseignement (UE) ».*
Contenu :	- 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 infos :	Ouvrages de référence 1. Berry M. and G. Linoff (2000), "Matering 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 Prdiction", 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 Prepation 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
Faculté ou entité en charge:	LSBA

Programmes / formations proposant cette unité d'enseignement (UE)

Intitulé du programme	Sigle	Crédits	Prérequis	Acquis d'apprentissage
Master [120] en linguistique	LING2M	5	-	
Certificat universitaire en statistique	STAT9CE	5	-	
Master [120] en statistiques, orientation générale	STAT2M	5	LSTAT2110 **et** LSTAT2120 **et** LSTAT2100	
Master [120] : ingénieur civil en mathématiques appliquées	MAP2M	4	-	
Master [120] en ingénieur de gestion	INGE2M	5	-	
Master [120] bioingénieur : chimie et bioindustries	BIRC2M	5	-	
Master [120] bioingénieur : sciences et technologies de l'environnement	BIRE2M	5	-	
Master [120] bioingénieur : gestion des forêts et des espaces naturels	BIRF2M	5	-	
Master [120] bioingénieur : sciences agronomiques	BIRA2M	5	-	
Certificat d'université : Statistique (15/30 crédits)	STAT2FC	5	-	