










4.00 crédits	15.0 h + 15.0 h	Q2
--------------	-----------------	----

Enseignants	Van Oirbeek Robin ;
Langue d'enseignement	Anglais
Lieu du cours	Louvain-la-Neuve
Préalables	Concepts et outils équivalents à ceux enseignés dans les UEs LSTAT2020 Logiciels et programmation statistique de base LSTAT2120 Linear models LSTAT2110 Analyse des données LSTAT2100 Modèles linéaires généralisés et données discrètes
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	<b>A la fin de cette unité d'enseignement, l'étudiant est capable de :</b>  A. Eu égard au référentiel AA du programme de master en statistique, orientation générale, cette activité contribue au développement et à l'acquisition des AA suivants : de manière prioritaire : 1.2, 1.3, 2.1, 2.2, 2.3, 2.5, 3.1, 3.3, 5.5, 6.3, 6.4 de manière secondaire : 2.4, 3.4, 5.1, 5.2  Eu égard au référentiel AA du programme de master en statistique, orientation biostatistique, cette activité contribue au développement et à l'acquisition des AA suivants : 1 de manière prioritaire : 1.2, 1.3, 2.1, 2.2, 2.3, 2.5, 3.1, 3.3, 6.1, 6.2. de manière secondaire : 2.4.  B. 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.
Contenu	<b>Introduction au data mining</b> <ul style="list-style-type: none"> <li>• Data et systèmes du data mining</li> <li>• Domaines d'application du data mining</li> <li>• Processus et méthodologie du data mining</li> <li>• Data mining dans le domaine du Customer Relationship Management (CRM)</li> <li>• Statistique classique et data mining</li> </ul> <b>Préparation des données pour data mining</b> <ul style="list-style-type: none"> <li>• Principales étapes dans la préparation des données</li> <li>• Spécification des données pour le data mining</li> <li>• Extraction et agrégation des données</li> <li>• Audit et exploration des données</li> <li>• Pré-traitement des données</li> </ul> <b>Modélisation prédictive</b> <ul style="list-style-type: none"> <li>• Arbres de décision</li> <li>• Réseaux de neurones</li> <li>• Validation et choix de modèles</li> </ul> <b>Modélisation descriptive</b> <ul style="list-style-type: none"> <li>• Clustering</li> <li>• K-means</li> <li>• Kohonen Self-Organising Map</li> </ul> <b>Etudes de cas</b>

<p>Bibliographie</p>	<ol style="list-style-type: none"> <li>1. Berry M. and G. Linoff (2000), "Matering Data Mining, The Art and Science of Customer Relationship Management", John Wiley.</li> <li>2. Bishop, C.M. (1995), Neural Networks for Pattern Recognition, Oxford.</li> <li>3. Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. (1984), "Classification and Regression Trees", Wadsworth, Inc., Belmont, California.</li> <li>4. Han J. and M. Kamber (2000), "Data Mining: Concepts and Techniques", Morgan Kaufmann,.</li> <li>5. Hastie Tr., R. Tibshirani and J. Friedman (2001), "The Elements of Statistical Learning -Data Mining, Inference and Prdiction", Springer.</li> <li>6. Haykin S., "Neural Networks: A comprehensive Foundation", Prentice Hall, 1999</li> <li>7. Kohonen T. (1995), "Self-Organizing Maps", Springer Series in Information Sciences, Oxford University Press.</li> <li>8. Piatetsky-Shapiro G. and W. J. Frawley (1991), "Knowledge Discovery in Databases", AAAI/MIT Press.</li> <li>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,.</li> <li>10. Pyle D. (2000), "Data Prepatation for Data Mining", Morgan Kaufman.</li> <li>11. Richard O. Dula, Pete E. Hart and David G. Stork (2000), "Pattern Classification", John Wiley, Second edition.</li> <li>12. Van Hulle M. (2000), "Faithful Representations and Topographic Maps: From Distortion- to Information-Based Self-Organization", John Willey</li> </ol>
<p>Faculté ou entité en charge:</p>	<p>LSBA</p>

Programmes / formations proposant cette unité d'enseignement (UE)				
Intitulé du programme	Sigle	Crédits	Prérequis	Acquis d'apprentissage
Master [120] en science des données, orientation statistique	DATS2M	5		
Master [120] en statistique, orientation biostatistiques	BSTA2M	4		
Master [120] en linguistique	LING2M	5		
Master [120] : bioingénieur en sciences et technologies de l'environnement	BIRE2M	5		
Master de spécialisation en méthodes quantitatives en sciences sociales	LMQS2MC	4		
Master [120] en sciences actuarielles	ACTU2M	4		
Master [120] en statistique, orientation générale	STAT2M	4		
Master [120] : bioingénieur en chimie et bioindustries	BIRC2M	5		
Master [120] : ingénieur civil en mathématiques appliquées	MAP2M	5		
Certificat d'université : Statistique et science des données (15/30 crédits)	STAT2FC	4		