









5.00 credits	15.0 h + 15.0 h	Q2
--------------	-----------------	----

Teacher(s)	Van Oirbeek Robin ;
Language :	English
Place of the course	Louvain-la-Neuve
Prerequisites	<i>The prerequisite(s) for this Teaching Unit (Unité d'enseignement – UE) for the programmes/courses that offer this Teaching Unit are specified at the end of this sheet.</i>
Main themes	- Data Mining application domains - Steps of a data mining project - Sampling and partitioning of the data base and training and validation sets - Data pretreatment and validation - Preliminary variable analysis, variables reduction and transformation - Classification and modeling tools of data mining - Decision trees - Neural networks - Tools to validate and compare estimated models - Case studies
Learning outcomes	<p><b>At the end of this learning unit, the student is able to :</b></p> <p>1 In this course, we will learn data mining methodology and techniques for knowledge discovery in large databases. We will also see how data mining differs from traditional statistics and how to treat a practical problem with an appropriate data mining tool.</p>
Content	<p><b>Introduction to data mining</b></p> <ul style="list-style-type: none"> <li>• Data and data mining systems</li> <li>• Data mining applications</li> <li>• Data mining process and methodology</li> <li>• Data mining in customer relationship management (CRM)</li> <li>• Traditional statistics versus data mining</li> </ul> <p><b>Data preparation for data mining</b></p> <ul style="list-style-type: none"> <li>• Data preparation stages</li> <li>• Data specification</li> <li>• Data extraction and aggregations</li> <li>• Data audit and exploration</li> <li>• Data pre-processing</li> </ul> <p><b>Predictive modelling</b></p> <ul style="list-style-type: none"> <li>• Decision trees</li> <li>• Neural networks</li> <li>• Model validation and assessment</li> </ul> <p><b>Descriptive modelling</b></p> <ul style="list-style-type: none"> <li>• Clustering</li> <li>• K-means</li> <li>• Kohonen Self-Organising Map</li> </ul> <p><b>Case studies</b></p>

Bibliography	<ol style="list-style-type: none"> <li>1. Berry M. and G. Linoff (2000), "Mastering 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 Prediction", 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 Preparation for Data Mining", Morgan Kaufman.</li> <li>11. Richard O. Duda, 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 Wiley</li> </ol>
Faculty or entity in charge	LSBA

Programmes containing this learning unit (UE)				
Program title	Acronym	Credits	Prerequisite	Learning outcomes
Master [120] in Linguistics	LING2M	5		
Master [120] in Statistics: General	STAT2M	5	LSTAT2110 AND LSTAT2120 AND LSTAT2100	
Master [120] in Environmental Bioengineering	BIRE2M	5		
Master [120] in Actuarial Science	ACTU2M	5		
Advanced Master in Quantitative Methods in the Social Sciences	LMQS2MC	5		
Master [120] in Chemistry and Bioindustries	BIRC2M	5		
Certificat d'université : Statistique et sciences des données (15/30 crédits)	STAT2FC	5		
Master [120] in Mathematical Engineering	MAP2M	5		
Master [120] in Data Science : Statistic	DATS2M	5		