










4.00 credits	15.0 h + 15.0 h	Q2
--------------	-----------------	----

Teacher(s)	Van Oirbeek Robin ;								
Language :	English								
Place of the course	Louvain-la-Neuve								
Prerequisites	<p>Concepts and tools equivalent to those taught in teaching units</p> <table border="0"> <tr> <td>LSTAT2020</td> <td>Logiciels et programmation statistique de base</td> </tr> <tr> <td>LSTAT2120</td> <td>Linear models</td> </tr> <tr> <td>LSTAT2110</td> <td>Analyse des données</td> </tr> <tr> <td>LSTAT2100</td> <td>Modèles linéaires généralisés et données discrètes</td> </tr> </table>	LSTAT2020	Logiciels et programmation statistique de base	LSTAT2120	Linear models	LSTAT2110	Analyse des données	LSTAT2100	Modèles linéaires généralisés et données discrètes
LSTAT2020	Logiciels et programmation statistique de base								
LSTAT2120	Linear models								
LSTAT2110	Analyse des données								
LSTAT2100	Modèles linéaires généralisés et données discrètes								
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>At the end of this learning unit, the student is able to :</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>Introduction to data mining</p> <ul style="list-style-type: none"> • Data and data mining systems • Data mining applications • Data mining process and methodology • Data mining in customer relationship management (CRM) • Traditional statistics versus data mining <p>Data preparation for data mining</p> <ul style="list-style-type: none"> • Data preparation stages • Data specification • Data extraction and aggregations • Data audit and exploration • Data pre-processing <p>Predictive modelling</p> <ul style="list-style-type: none"> • Decision trees • Neural networks • Model validation and assessment <p>Descriptive modelling</p> <ul style="list-style-type: none"> • Clustering • K-means • Kohonen Self-Organising Map <p>Case studies</p>								

Bibliography	<ol style="list-style-type: none"> 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. Duda, 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
Faculty or entity in charge	LSBA

Programmes containing this learning unit (UE)				
Program title	Acronym	Credits	Prerequisite	Learning outcomes
Master [120] in Data Science : Statistic	DATS2M	5		
Master [120] in Statistics: Biostatistics	BSTA2M	4		
Master [120] in Linguistics	LING2M	5		
Master [120] in Environmental Bioengineering	BIRE2M	5		
Advanced Master in Quantitative Methods in the Social Sciences	LMQS2MC	5		
Master [120] in Actuarial Science	ACTU2M	4		
Master [120] in Statistics: General	STAT2M	4		
Master [120] in Chemistry and Bioindustries	BIRC2M	5		
Master [120] in Mathematical Engineering	MAP2M	5		
Certificat d'université : Statistique et science des données (15/30 crédits)	STAT2FC	4		