

5.0 credits	15.0 h + 15.0 h	2q
-------------	-----------------	----

Teacher(s) :	Chen Libei ;
Language :	Français
Place of the course	Louvain-la-Neuve
Main themes :	<ul style="list-style-type: none"> - 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
Aims :	<p>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> <p><i>The contribution of this Teaching Unit to the development and command of the skills and learning outcomes of the programme(s) can be accessed at the end of this sheet, in the section entitled "Programmes/courses offering this Teaching Unit".</i></p>
Content :	<ul style="list-style-type: none"> - Introduction to data mining <ul style="list-style-type: none"> o Data and data mining systems o Data mining applications o Data mining process and methodology o Data mining in customer relationship management (CRM) o Traditional statistics versus data mining - Data preparation for data mining <ul style="list-style-type: none"> o Data preparation stages o Data specification o Data extraction and aggregations o Data audit and exploration o Data pre-processing - Predictive modelling <ul style="list-style-type: none"> o Decision trees o Neural networks o Model validation and assessment - Descriptive modelling <ul style="list-style-type: none"> o Clustering o K-means o Kohonen Self-Organising Map - Case studies

<p>Other infos :</p>	<p>References:</p> <ol style="list-style-type: none"> 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 Preperation 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 & Sons Inc.
<p>Cycle and year of study :</p>	<p>> Master [120] in Mathematical Engineering > Master [120] in Linguistics > Certificat universitaire en statistique > Master [120] in Statistics: General</p>
<p>Faculty or entity in charge:</p>	<p>LSBA</p>