UCLouvain

ldats2350

2020

Data Mining

Due to the COVID-19 crisis, the information below is subject to change, in particular that concerning the teaching mode (presential, distance or in a comodal or hybrid format).

5 credits	15.0 h + 15.0 h	Q2

Teacher(s)	Van Oirbeek Robin ;
Language :	English
Place of the course	Louvain-la-Neuve
Prerequisites	*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.*
Main themes	- Data Mining application domains - Steps of a data mining project - Sampling and partionning of the data base and training and validation sets - Data pretreatment and validation - Premilinary 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	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. - - - - *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".*
Content	**Introduction to data mining** • 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 **Data preparation for data mining** • Data preparation stages • Data specification • Data extraction and aggregations • Data audit and exploration • Data pre-processing **Predictive modelling** • Decision trees • Neural networks • Model validation and assessment **Descriptive modelling** • Clustering • K-means • Kohonen Self-Organising Map **Case studies**

Bibliography	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
Faculty or entity in charge	LSBA

Programmes containing this learning unit (UE)

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