

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

30.0 h + 15.0 h






Q2

Teacher(s)	Saerens Marco ;
Language :	English
Place of the course	Louvain-la-Neuve
Main themes	<p>The course is structured around four themes</p> <ol style="list-style-type: none"> <li>1. Complements of data mining,</li> <li>2. Decision making,</li> <li>3. Information retrieval,</li> <li>4. Link analysis and web/graph mining .</li> </ol>
Aims	<p>Given the learning outcomes of the "Master in Computer Science and Engineering" program, this course contributes to the development, acquisition and evaluation of the following learning outcomes:</p> <ul style="list-style-type: none"> <li>• INFO1.1-3</li> <li>• INFO2.2-3</li> <li>• INFO5.2</li> </ul> <p>Given the learning outcomes of the "Master [120] in Computer Science" program, this course contributes to the development, acquisition and evaluation of the following learning outcomes:</p> <ol style="list-style-type: none"> <li>1 • SINF1.M4</li> <li>• SINF2.2-3</li> <li>• SINF5.2</li> </ol> <p>Students completing this course successfully will be able to</p> <ul style="list-style-type: none"> <li>• explain quantitative and qualitative data mining methods and to apply them to decision making</li> <li>• develop a critical view of data mining techniques in specific application domains</li> <li>• master information retrieval techniques from very large data collection, possibly enriched with link structures (WEB, social networks, ...)</li> <li>• explain application of information retrieval techniques in the context of search engines and automated recommendation systems</li> <li>• implement data mining and information retrieval algorithms within standard software environments such as S-Plus, R, SAS, Weka or Matlab</li> </ul> <p>-----</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>
Evaluation methods	<p><b>Due to the COVID-19 crisis, the information in this section is particularly likely to change.</b></p> <ul style="list-style-type: none"> <li>• One or two projects for 6 points on 20 to 10 points on 20 (for both projects), depending on the size and the number of these projects. This will be specified at the first or second lecture.</li> <li>• Oral or written exam (depending on the health situation and the number of students) : 14/20 to 10/20 (depending on the scenario concerning the projects).</li> </ul>
Content	<p>The content changes from year to year, but the chapters with a * are always taught.</p> <p>* Complements of data mining</p> <ul style="list-style-type: none"> <li>• Principal components analysis</li> <li>• Canonical correlation analysis</li> <li>• Correspondence analysis</li> <li>• Log-linear models</li> <li>• Discriminant analysis</li> <li>• Multidimensional scaling</li> <li>• Markov and hidden Markov models</li> <li>• etc</li> </ul> <p>* Decision making</p>

	<ul style="list-style-type: none"> <li>• Dynamic programming and applications</li> <li>• Markov decision processes and reinforcement learning</li> <li>• Exploration/exploitation and bandit problems</li> <li>• Utility theory</li> <li>• Multi-criteria preference modeling - the Promethee method</li> <li>• Probabilistic reasoning with bayesian networks</li> <li>• Two-players game theory</li> <li>• Collective decisions</li> </ul> <p>* Information retrieval</p> <ul style="list-style-type: none"> <li>• The basic vector-space model</li> <li>• The probabilistic model</li> <li>• Ranking web pages : PageRank, HITS, etc.</li> <li>• Collaborative recommendation models (recommender systems) .</li> </ul> <p>Link analysis and web/graph mining</p> <ul style="list-style-type: none"> <li>• Network community detection</li> <li>• Similarity measures between nodes</li> <li>• Spectral graph partitioning and mapping</li> </ul> <p>* Reputation and collaborative recommendation models</p>
Inline resources	Available on Moodle
Bibliography	<p>Some recommended reference books :</p> <ul style="list-style-type: none"> <li>• Alpaydin (2004), "Introduction to machine learning". MIT Press.</li> <li>• Bardos (2001), "Analyse discriminante. Application au risque et scoring financier. Dunod.</li> <li>• Bishop (1995), "Neural networks for pattern recognition". Clarendon Press.</li> <li>• Bishop (2006), "Pattern recognition and machine learning". Springer-Verlag.</li> <li>• Bourouche &amp; Saporta (1983), "L'analyse des données". Que Sais-je.</li> <li>• Cornuéjols &amp; Miclet (2002), "Apprentissage artificiel. Concepts et algorithmes". Eyrolles.</li> <li>• Duda, Hart &amp; Stork (2001), "Pattern classification, 2nd ed". John Wiley &amp; Sons.</li> <li>• Dunham (2003), "Data mining. Introductory and advanced topics". Prentice-Hall.</li> <li>• Greenacre (1984), "Theory and applications of correspondence analysis". Academic Press.</li> <li>• Han &amp; Kamber (2005), "Data mining: Concepts and techniques, 2nd ed.". Morgan Kaufmann.</li> <li>• Hand (1981), "Discrimination and classification". John Wiley &amp; Sons.</li> <li>• Hardle &amp; Simar (2003), "Applied multivariate statistical analysis". Springer-Verlag. Disponible à <a href="http://www.quantlet.com/mdstat/scripts/mva/htmlbook/mvahtml.html">http://www.quantlet.com/mdstat/scripts/mva/htmlbook/mvahtml.html</a></li> <li>• Hastie, Tibshirani &amp; Friedman (2001), "The elements of statistical learning". Springer-Verlag.</li> <li>• Johnson &amp; Wichern (2002), "Applied multivariate statistical analysis, 5th ed". Prentice-Hall.</li> <li>• Lebart, Morineau &amp; Piron (1995), "Statistique exploratoire multidimensionnelle". Dunod.</li> <li>• Mitchell (1997), "Machine learning". McGraw-Hill.</li> <li>• Naim, Wuillemin, Leray, Pourret &amp; Becker (2004), "Réseaux bayesiens". Editions Eyrolles.</li> <li>• Nilsson (1998), "Artificial intelligence: A new synthesis". Morgan Kaufmann.</li> <li>• Ripley (1996), "Pattern recognition and neural networks". Cambridge University Press.</li> <li>• Rosner (1995), "Fundamentals of biostatistics, 4th ed". Wadsworth Publishing Company.</li> <li>• Saporta (1990), "Probabilités, analyse des données et statistique". Editions Technip.</li> <li>• Tan, Steinbach &amp; Kumer (2005), "Introduction to data mining". Pearson.</li> <li>• Theodoridis &amp; Koutroumbas (2003), "Pattern recognition, 3th ed". Academic Press.</li> <li>• Therrien (1989), "Decision, estimation and classification". Wiley &amp; Sons.</li> <li>• Venables &amp; Ripley (2002), "Modern applied statistics with S. Springer-Verlag.</li> <li>• Webb (2002), "Statistical pattern recognition, 2nd ed". John Wiley and Sons.</li> </ul>
Other infos	<p>Background :</p> <ul style="list-style-type: none"> <li>• LBIR1304 ou LFSAB1105 : a course on probability theory and mathematical statistics,</li> <li>• LBIR1200 ou LFSAB1101 : an undergraduate course on matrix algebra,</li> <li>• LFSAB1402 : a course on the basis of programming</li> </ul>
Faculty or entity in charge	INFO

### Force majeure

Evaluation methods	<p>Provided the health situation, it could be that the final exam has to be organized remotely and that its form changes from one exam session to another, during the same academic year. In that case, the method of assessment will take the form of a written or a multiple-choice exam (possibly with proctoring via Teams or another software), or even a remote oral exam if the number of students is not too large. These modalities will be communicated to the students within the official time frame provided by the university.</p>
--------------------	--

Programmes containing this learning unit (UE)				
Program title	Acronym	Credits	Prerequisite	Aims
Master [120] in Data Science : Statistic	DATS2M	5		
Master [120] in Computer Science and Engineering	INFO2M	5		
Master [120] in Computer Science	SINF2M	5		
Certificat d'université : Statistique et sciences des données (15/30 crédits)	STAT2FC	5		
Master [120] in Actuarial Science	ACTU2M	5		
Master [120] in Mathematical Engineering	MAP2M	5		
Master [120] in Agricultural Bioengineering	BIRA2M	5		
Master [120] in Forests and Natural Areas Engineering	BIRF2M	5		
Master [120] in Environmental Bioengineering	BIRE2M	5		
Master [120] in Data Science Engineering	DATE2M	5		
Master [120] in Chemistry and Bioindustries	BIRC2M	5		
Master [120] in Data Science: Information Technology	DATI2M	5		
Master [120] in Statistic: General	STAT2M	5		