UCLouvain

linfo2275

2025

Data mining & decision making

5.00 credits	30.0 h + 15.0 h	Q2

Teacher(s)	Saerens Marco ;
Language :	English > French-friendly
Place of the course	Louvain-la-Neuve
Main themes	The course is structured around four themes 1. Complements of data mining, 2. Decision making, 3. Information retrieval, 4. Link analysis and web/graph mining .
Learning outcomes	**At the end of this learning unit, the student is able to :** 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: · INFO1.1-3 · INFO2.2-3 · INFO5.2 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: · SINF1.M4 · SINF2.2-3 · SINF5.2 Students completing this course successfully will be able to · explain quantitative and qualitative data mining methods and to apply them to decision making; · develop a critical view of data mining techniques in specific application domains; · master information retrieval techniques from very large data collection, possibly enriched with link structures (WEB, social networks, ...); · explain application of information retrieval techniques in the context of search engines and automated recommendation systems; · implement data mining and information retrieval algorithms within standard software environments such as S-Plus, R, SAS, Weka or Matlab.
Evaluation methods	· 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. If you failed the project in June, you are allowed to redo and resubmit it for the August session before the first day of the exam session, 23:55 (send it to the professor and the assistant). If you do not resubmit the project in August, we take back the mark obtained in June (0 if you never submitted it). · Oral or written exam (depending on the situation and the number of students) : 14/20 to 10/20 (depending on the scenario concerning the projects). · The exam is **mandatory** , also in August – you are considered as absent if you do not pass it. Concernant le projet/cas d'étude obligatoire et l'utilisation d'IA de type Chat GPT, assurez-vous que : "En soumettant un travail pour évaluation, vous affirmez : (i) qu'il reflète fidèlement le phénomène étudié, et pour cela vous devez avoir vérifié les faits, surtout s'ils sont prétendus par une IA générative (dont vous devez mentionner explicitement l'utilisation en tant qu'outil de soutien à la réalisation de votre travail) ; (ii) avoir respecté toutes les exigences spécifiques du travail qui vous est confié, notamment les exigences pour la transparence et la documentation de la démarche scientifique mise en œuvre. Si l'une de ces affirmations n'est pas vraie, que ce soit intentionnellement ou par négligence, vous êtes en défaut de votre engagement déontologique vis-à-vis de la connaissance produite dans le cadre de votre travail, et éventuellement d'autres aspects de l'intégrité académique, ce qui constitue une faute académique et sera considéré comme tel".
Content	The content changes from year to year, but the chapters with a * are always teached. The teached sections depend on the year. * Complements of data mining · Principal components analysis · Canonical correlation analysis

	· Correspondence analysis · Log-linear models · Discriminant analysis · Multidimensional scaling · Markov and hidden Markov models · etc * Decision making · Dynamic programming and applications · Markov decision processes and reinforcement learning · Exploration/exploitation and bandit problems · Utility theory · Multi-criteria preference modeling - the Promethee method · Probabilistic reasoning with bayesian networks · Two-players game theory · Collective decisions * Information retrieval · The basic vector-space model · The probabilistic model · Ranking web pages : PageRank, HITS, etc. · Collaborative recommendation models (recommender systems) . Link analysis and web/graph mining · Network community detection · Similarity measures between nodes · Spectral graph partitioning and mapping * Reputation and collaborative recommendation models
Inline resources	Available on Moodle
Bibliography	Some recommended reference books : · Alpaydin (2004), "Introduction to machine learning". MIT Press. · Bardos (2001), "Analyse discriminante. Application au risque et scoring financier. Dunod. · Bishop (1995), "Neural networks for pattern recognition". Clarendon Press. · Bishop (2006), "Pattern recognition and machine learning". Springer-Verlag. · Bouroche & Saporta (1983), "L'analyse des données". Que Sais-je. · Cornuéjols & Miclet (2002), "Apprentissage artificiel. Concepts et algorithmes". Eyrolles. · Duda, Hart & Stork (2001), "Pattern classification, 2nd ed". John Wiley & Sons. · Dunham (2003), "Data mining. Introductory and advanced topics". Prentice-Hall. · Greenacre (1984), "Theory and applications of correspondence analysis". Academic Press. · Han & Kamber (2005), "Data mining: Concepts and techniques, 2nd ed.". Morgan Kaufmann. · Hand (1981), "Discrimination and classification". John Wiley & Sons. · Hardle & Simar (2003), "Applied multivariate statistical analysis". Springer-Verlag. Disponible à http://www.quantlet.com/mdstat/scripts/mva/htmlbook/mvahtml.html · Hastie, Tibshirani & Friedman (2001), "The elements of statistical learning". Springer-Verlag. · Johnson & Wichern (2002), "Applied multivariate statistical analysis, 5th ed". Prentice-Hall. · Lebart, Morineau & Piron (1995), "Statistique exploratoire multidimensionnelle". Dunod. · Mitchell (1997), "Machine learning". McGraw-Hill. · Naim, Wuillemin, Leray, Pourret & Becker (2004), "Réseaux bayesiens". Editions Eyrolles. · Nilsson (1998), "Artificial intelligence: A new synthesis". Morgan Kaufmann. · Ripley (1996), "Pattern recognition and neural networks". Cambridge University Press. · Rosner (1995), "Fundamentals of biostatistics, 4th ed".Wadsworth Publishing Company. · Saporta (1990), "Probabilités, analyse des données et statistique". Editions Technip. · Tan, Steinbach & Kumer (2005), "Introduction to data mining". Pearson. · Theodoridis & Koutroumbas (2003), "Pattern recognition, 3th ed". Academic Press. · Therrien (1989), "Decision, estimation and classification". Wiley & Sons. · Venables & Ripley (2002), "Modern applied statistics with S. Springer-Verlag. · Webb (2002), "Statistical pattern recognition, 2nd ed". John Wiley and Sons.
Other infos	Background / prerequisites : · LBIR1304 ou LFSAB1105 : a course on probability theory and mathematical statistics, · LBIR1200 ou LFSAB1101 : a course on linear and matrix algebra, · LFSAB1402 : a good Python programming course, · A course in multivariate calculus (mathematics).
Faculty or entity in charge	INFO

Programmes containing this learning unit (UE)

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