

5.00 crédits

30.0 h + 15.0 h

Q2

Enseignants	Saerens Marco ;
Langue d'enseignement	Anglais > Facilités pour suivre le cours en français
Lieu du cours	Louvain-la-Neuve
Thèmes abordés	Le cours est articulé autour de quatre thèmes, 1. Compléments de fouille de données 2. Prise de décision , 3. Recherche d'information, 4. A nalyse de liens et l'exploration du web / graphique .
Acquis d'apprentissage	<p>A la fin de cette unité d'enseignement, l'étudiant est capable de :</p> <p>Eu égard au référentiel AA du programme « Master ingénieur civil en informatique », ce cours contribue au développement, à l'acquisition et à l'évaluation des acquis d'apprentissage suivants :</p> <ul style="list-style-type: none"> • INFO1.1-3 • INFO2.2-3 • INFO5.2 <p>Eu égard au référentiel AA du programme « Master [120] en sciences informatiques », ce cours contribue au développement, à l'acquisition et à l'évaluation des acquis d'apprentissage suivants :</p> <ul style="list-style-type: none"> • SINF1.M4 • SINF2.2-3 • SINF5.2 <p>1 Les étudiants ayant suivi avec fruit ce cours seront capables de</p> <ul style="list-style-type: none"> • expliquer et appliquer des méthodes de fouille de données (data mining), tant qualitatives que quantitatives, dans le cadre de la prise de décision. • avoir un jugement critique sur les méthodes de fouilles de données en fonction de leur champ d'application. • maîtriser des méthodes d'extraction et recherche d'informations (information retrieval) dans de très larges collections de données, éventuellement enrichies d'une structure de liens (WEB, réseaux sociaux...). • expliquer l'application de ces méthodes dans les moteurs de recherche et les systèmes de recommandation automatisée. • mettre en oeuvre les algorithmes de fouille de données et d'extraction d'informations dans des logiciels de data mining ou de traitement statistique tels que S-Plus, R, SAS, Weka ou Matlab.
Modes d'évaluation des acquis des étudiants	<ul style="list-style-type: none"> • Un seul projet important ou deux petits projets en cours d'année comptant au total pour 6/20 à 10/20, selon l'ampleur et le nombre de ces projets (déterminé en début d'année académique, en leçon 1 ou 2). • Examen oral ou écrit (selon la situation et le nombre d'étudiants): noté sur 14/20 à 10/20, selon le scénario du projet ci-dessus. • L'examen est obligatoire, y compris en août. Vous êtes considéré comme absent si vous ne le passez pas. <p>Concernant le projet/cas d'étude obligatoire et l'utilisation d'IA de type Chat GPT, assurez-vous que :</p> <p>"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".</p>
Contenu	<p>Le contenu varie d'année en année, mais les chapitres précédés d'une * sont toujours enseignés. Les sections enseignées varient d'une année à l'autre.</p> <p>* Compléments de fouille de données (méthodes statistiques multivariées et données séquentielles)</p> <ul style="list-style-type: none"> • Analyse en composantes principales • Analyse des corrélations canoniques

	<ul style="list-style-type: none"> • Analyse des correspondances • Modèles log-linéaires • Analyse discriminante • Multidimensional scaling • Modèles de Markov et modèles de Markov cachés • etc <p>* Prise de décision</p> <ul style="list-style-type: none"> • Programmation dynamique et applications • Processus de décision de Markov et l'apprentissage par renforcement • Exploration / exploitation • Théorie de l'utilité • Modélisation des préférences multi-critères - la méthode Prométhée • Raisonnement probabiliste avec des réseaux bayésiens • Théorie des possibilités • Théorie des jeux à deux joueurs • Décisions collectives <p>* Recherche d'information</p> <ul style="list-style-type: none"> • Modèle de base d'espace vectoriel • Modèle probabiliste • Pages Web Ranking: PageRank, HITS, etc. • Modèles collaboratifs par recommandations (systèmes de recommandation). <p>* Analyse de liens et l'exploration du web / réseaux</p> <ul style="list-style-type: none"> • Détection de la communauté réseau • Mesures de similarité entre les noeuds • Partitionnement de graphe spectral et cartographie <p>* Modèles de réputation et de recommandation collaborative</p>
Ressources en ligne	Disponibles sur la page Moodle du cours
Bibliographie	<p>Some recommended reference books :</p> <ul style="list-style-type: none"> • 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 bayésiens". 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.
Autres infos	<p>Préalables :</p> <ul style="list-style-type: none"> • LBIR1304 ou LFSAB1105 : un cours de calcul des probabilités et de statistique mathématique, • LBIR1200 ou LFSAB1101 : un cours de calcul matriciel et d'algèbre linéaire, • LFSAB1402 : un bon cours de programmation Python, • Un cours d'analyse multivariée (mathématiques).
Faculté ou entité en charge:	INFO

Programmes / formations proposant cette unité d'enseignement (UE)				
Intitulé du programme	Sigle	Crédits	Prérequis	Acquis d'apprentissage
Master [120] en science des données, orientation statistique	DATS2M	5		
Master [120] : bioingénieur en gestion des forêts et des espaces naturels	BIRF2M	5		
Master [120] : bioingénieur en sciences et technologies de l'environnement	BIRE2M	5		
Master [120] en sciences actuarielles	ACTU2M	5		
Master [120] : bioingénieur en chimie et bioindustries	BIRC2M	5		
Master [120] : ingénieur civil en informatique	INFO2M	5		
Master [120] en sciences informatiques	SINF2M	5		
Master [120] : ingénieur civil en mathématiques appliquées	MAP2M	5		
Master [120] : ingénieur civil en science des données	DATE2M	5		
Certificat d'université : Statistique et science des données (15/30 crédits)	STAT2FC	5		
Master [120] : bioingénieur en sciences agronomiques	BIRA2M	5		
Master [120] en science des données, orientation technologies de l'information	DATI2M	5		