

Données Intelligentes <i>Smart Data</i>			
Code cours <i>Course code: DOI</i>		Crédits ECTS <i>ECTS Credits: 2</i>	
Département <i>Department</i>	: IA	Cours Lectures	: 15H
Coordonnateurs <i>Lecturers</i>	: CHARDIN Brice	T.D. Tutorials	: 15H
Période <i>Year of study</i>	: 3 ^{ème} année <i>3rd year</i>	T.P. Laboratory sessions	:
Semestre <i>Semester</i>	: 5 ^{ème} semestre – B <i>5th semester - B</i>	Projet Project	:
Evaluation <i>Assessment method(s)</i>	: 1 examen écrit, <i>1 written exam</i>	Non encadré Homework	:
Langue d’instruction <i>Language of instruction</i>	: Français <i>French</i>	Horaire global Total hours	: 30H
Type de cours <i>Type of course</i>	: Obligatoire <i>Compulsory</i>		
Niveau <i>Level of course</i>	: Avancé <i>graduate</i>		

Compétences attendues :

- Modéliser un problème de décision en termes de programme d'optimisation linéaire.
- Savoir utiliser un solveur et interpréter les résultats.
- Comprendre l'importance du prétraitement des données et connaître les techniques applicables.
- Comprendre et savoir utiliser les techniques principales d'extraction de connaissances et d'apprentissage automatique.

Pré-requis :

- Utilisation et exploitation des données (UED1)
- Programmation, modélisation et algorithmique (INF1 et INF2)
- Analyse numérique (IFM2, CSA3 et CSB3)

Contenu :

- Programmation par contrainte
 - programmation linéaire
 - programmation linéaire en nombre entiers
 - problèmes de satisfaction de contraintes
 - processus de modélisation, résolution et interprétation
 - implémentation avec IBM ILOG CPLEX
- Prétraitement des données
 - impact de la qualité des données sur les techniques d'apprentissage automatique
 - identification et correction des données erronées
 - processus ETL (Extract, Transform, Load)
- Fouille de données
 - itemsets fréquents
 - règles d'association
 - dépendances fonctionnelles et variantes
 - mesures d'intérêt applicables aux règles générées
- Apprentissage automatique
 - problèmes de classification et de régression
 - classification non supervisée (partitionnement)
 - apprentissage supervisé (arbres et forêts de décision, SVM, réseaux de neurones)
 - méthodologie de validation expérimentale

Bibliographie :

- Dantzig George, Thapa Mukund, Linear Programming 1: Introduction. Springer Series in Operations Research and Financial Engineering, 2003.
- Dantzig George, Thapa Mukund, Linear Programming 2: Theory and Extensions. Springer Series in Operations Research and Financial Engineering, 2003.
- Jiawei Han, Micheline Kamber, Jian Pei. 2011. Data Mining: Concepts and Techniques. 3rd Edition. ISBN: 978-0-12-381479-1. Elsevier, 2011.
- Marriott , K., Stuckey, P., Programming with constraints : an introduction, The MIT Press, 1998
- Mazhar Hameed and Felix Naumann. 2020. Data Preparation: A Survey of Commercial Tools. SIGMOD Rec. 49, 3 (September 2020), 18–29.
- Panik, M.J., Linear Programming: Mathematics, Theory and Algorithms, 1996.
- Vasilyev, F.P., Ivanitskiy, A.Y., In-Depth Analysis of Linear Programming, 2001

- Wolsey, L., Integer programming, Wiley-Interscience series in discrete mathematics and optimization. J., Wiley & sons, New York (N.Y.), Chichester, Weinheim, 1998

Expected competencies:

- Model a decision problem as a mathematical optimization program.
- Use a linear programming solver and interpret its results.
- Grasp the importance of data preprocessing and know its main associated techniques.
- Understand and use major data mining and machine learning techniques.

Prerequisites:

- Introduction to data management (UED1)
- Computer programming, data structures and algorithms (INF1 and INF2)
- Numerical analysis (IFM2, CSA3 and CSB3)

Content:

- Constraint programming
 - linear programming
 - integer linear programming
 - constraint satisfaction problems
 - model formulation, solving and results interpretation
 - implementation with IBM ILOG CPLEX
- Data preprocessing
 - impact of data quality on machine learning techniques
 - data cleaning
 - ETL process (Extract, Transform, Load)
- Data mining
 - frequents itemsets
 - association rule mining
 - functional dependencies and other rules
 - rule quality measures
- Machine learning
 - classification and regression problems
 - clustering
 - supervised learning (decision trees, random forests, SVM, neural networks)
 - experimental validation strategies

Recommended reading:

- Dantzig George, Thapa Mukund, Linear Programming 1: Introduction. Springer Series in Operations Research and Financial Engineering, 2003.
- Dantzig George, Thapa Mukund, Linear Programming 2: Theory and Extensions. Springer Series in Operations Research and Financial Engineering, 2003.
- Jiawei Han, Micheline Kamber, Jian Pei. 2011. Data Mining: Concepts and Techniques. 3rd Edition. ISBN: 978-0-12-381479-1. Elsevier, 2011.
- Marriott , K., Stuckey, P., Programming with constraints : an introduction, The MIT Press, 1998
- Mazhar Hameed and Felix Naumann. 2020. Data Preparation: A Survey of Commercial Tools. SIGMOD Rec. 49, 3 (September 2020), 18–29.
- Panik, M.J., Linear Programming: Mathematics, Theory and Algorithms, 1996.
- Vasilyev, F.P., Ivanitskiy, A.Y., In-Depth Analysis of Linear Programming, 2001
- Wolsey, L., Integer programming, Wiley-Interscience series in discrete mathematics and optimization. J., Wiley & sons, New York (N.Y.), Chichester, Weinheim, 1998