Master 2 M2 Data Science (DS)

Année universitaire 2023-2024

Information générale

Objectifs	
Responsable(s)	LE CAPITAINE HOEL
Mention(s) incluant ce parcours	master Informatique
Lieu d'enseignement	Polytech Nantes (site de la Chantrerie)
Langues / mobilité internationale	Parcours intégralement enseigné en anglais (proposant un enseignement de FLE en complément sur le site de Polytech Nantes).
Stage / alternance	
Poursuite d'études /débouchés	
Autres renseignements	
Conditions d'obtention de l'année	L'année est validée si la partie théorique est validée en première ou deuxième session (moyenne supérieure ou égale à 10/20) et si l'UE correspondant au stage est également validée avec une note supérieure ou égale à 10/20.

Programme

1er SEMESTRE	Code	ECTS	CM	CM (P)	CM (DS)	CM (DA)	CI	CI (P)	CI (DS)	CI (DA)	TD	TD (P)	TD (DS)	TD (DA)	TP	TP (P)	TP (DS)	TP (DA)	Distanciel	Total
Groupe d'UE : Semestre 3 de Data Science (30 ECTS)																				
Research methodology and case study		2	0	0	0	0	15	0	0	0	0	0	0	0	0	0	0	0	2.5	17.5
Data economics, law and ethics		3	0	0	0	0	21.5	0	0	0	0	0	0	0	0	0	0	0	2.5	24
Data dependencies and data integration		3	0	0	0	0	14	0	0	0	0	0	0	0	7.5	0	0	0	0	21.5
Visual analytics		3	0	0	0	0	14	0	0	0	0	0	0	0	7.5	0	0	0	2.5	24
Pattern mining and social network analysis		3	0	0	0	0	14	0	0	0	0	0	0	0	7.5	0	0	0	2.5	24
Text and sequential pattern mining		3	0	0	0	0	14	0	0	0	0	0	0	0	7.5	0	0	0	2.5	24
Cluster analysis and indexing		3	0	0	0	0	14	0	0	0	0	0	0	0	7.5	0	0	0	2.5	24
Classification, representation learning and dimensionality reduction		3	0	0	0	0	14	0	0	0	0	0	0	0	7.5	0	0	0	2.5	24
Probabilistic Graphical Models and statistical relational learning		3	0	0	0	0	14	0	0	0	0	0	0	0	7.5	0	0	0	2.5	24
Semantic knowledge representation		3	0	0	0	0	14	0	0	0	0	0	0	0	7.5	0	0	0	2.5	24
Conferences and invited courses (DS)		1	0	0	0	0	21.5	0	0	0	0	0	0	0	0	0	0	0	2.5	24
Groupe d'UE : Option pour public non francophone (sous conditions) (0 ECTS)																				
French langage and European culture		0	0	0	0	0	0	0	0	0	48	0	0	0	0	0	0	0	0	48
	Total	30																	25.00	**303.00**

2ème SEMESTRE	Code	ECTS	CM	CM (P)	CM (DS)	CM (DA)	CI	CI (P)	CI (DS)	CI (DA)	TD	TD (P)	TD (DS)	TD (DA)	TP	TP (P)	TP (DS)	TP (DA)	Distanciel	Total
Groupe d'UE : Semestre 4 de Data Science (30 ECTS)																				
Internship (DS)		30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Total	30																	0.00	**0.00**

Modalités d’évaluation

Mention Master 2ème année | Parcours : M2 Data Science (DS) | Année universitaire 2023-2024

Responsable(s) : LE CAPITAINE HOEL

REGIME ORDINAIRE

					PREMIERE SESSION							DEUXIEME SESSION							TOTAL	
					Contrôle continu			Examen				Contrôle continu			Examen				Coeff.	ECTS
	CODE UE	INTITULE	UE non dipl.		écrit	prat.	oral	écrit	prat.	oral	durée	ecrit	prat.	oral	écrit	prat.	oral	durée		
Groupe d'UE : Semestre 3 de Data Science																				
3		Research methodology and case study	N	obligatoire	1.4		0.6					1.4		0.6					2	2
3		Semantic knowledge representation	N	obligatoire		1.5		1.5					1.5		1.5				3	3
3		Probabilistic Graphical Models and statistical relational learning	N	obligatoire		0.9		2.1					0.9		2.1				3	3
3		Classification, representation learning and dimensionality reduction	N	obligatoire		1.2		1.8					1.2		1.8				3	3
3		Cluster analysis and indexing	N	obligatoire		1.2		1.8					1.2		1.8				3	3
3		Text and sequential pattern mining	N	obligatoire		1.2		1.8					1.2		1.8				3	3
3		Pattern mining and social network analysis	N	obligatoire		1.5		1.5					1.5		1.5				3	3
3		Visual analytics	N	obligatoire	0.45	0.9	0.45	1.2				0.45	0.9	0.45	1.2				3	3
3		Data dependencies and data integration	N	obligatoire		1.2		1.8					1.2		1.8				3	3
3		Data economics, law and ethics	N	obligatoire		1.5		1.5					1.5		1.5				3	3
3		Conferences and invited courses (DS)	N	obligatoire		1							1						1	1
Groupe d'UE : Option pour public non francophone (sous conditions)																				
3		French langage and European culture	N	optionnelle															0	0
Groupe d'UE : Semestre 4 de Data Science																				
4		Internship (DS)	N	obligatoire	7.5	15	7.5					7.5	15	7.5					30	30
																		TOTAL	60	60

A la seconde session, les notes de contrôle continu correspondent à un report des notes de CC de la première session.

DISPENSE D'ASSIDUITE

					PREMIERE SESSION							DEUXIEME SESSION							TOTAL	
					Contrôle continu			Examen				Contrôle continu			Examen					
	CODE UE	**INTITULE**	**UE non dipl.**		écrit	prat.	oral	écrit	prat.	oral	durée	ecrit	prat.	oral	écrit	prat.	oral	durée	**Coeff.**	**ECTS**
Groupe d'UE : Semestre 3 de Data Science																				
3		Research methodology and case study	N	obligatoire					2							2			2	2
3		Semantic knowledge representation	N	obligatoire				3							3				3	3
3		Probabilistic Graphical Models and statistical relational learning	N	obligatoire				3							3				3	3
3		Classification, representation learning and dimensionality reduction	N	obligatoire				3							3				3	3
3		Cluster analysis and indexing	N	obligatoire				3							3				3	3
3		Text and sequential pattern mining	N	obligatoire				1.8	1.2						1.8	1.2			3	3
3		Pattern mining and social network analysis	N	obligatoire				3							3				3	3
3		Visual analytics	N	obligatoire				3							3				3	3
3		Data dependencies and data integration	N	obligatoire				3							3				3	3
3		Data economics, law and ethics	N	obligatoire				3							3				3	3
3		Conferences and invited courses (DS)	N	obligatoire					1							1			1	1
Groupe d'UE : Option pour public non francophone (sous conditions)																				
3		French langage and European culture	N	optionnelle															0	0
Groupe d'UE : Semestre 4 de Data Science																				
4		Internship (DS)	N	obligatoire															30	30
																		TOTAL	60	60

A la seconde session, les notes de contrôle continu correspondent à un report des notes de CC de la première session.

Description des UE

	Research methodology and case study
Lieu d'enseignement	Polytech Nantes
Niveau	Master
Semestre	3
Responsable de l'UE	
Volume horaire total	**TOTAL :** 17.5h Répartition : **CM :** 0h **TD :** 0h **CI :** 15h **TP :** 0h **EAD :** 2.5h
Place de l'enseignement	
UE pré-requise(s)	
Parcours d'études comprenant l'UE	M2 Data Science (DS)
Evaluation	
Pondération pour chaque matière	Research methodology and case study **100%**
Obtention de l'UE	
Programme	
Objectifs (résultats d'apprentissage)	Upon completion, the student will have gained experience in semi-autonomous research work, yet with a guided methodology. They will get personalised tutoring on how to apply the general guidelines of the teaching unit to their work.
Contenu	Professional insertion: research & development processes in company processes. Goals and organisation of the scientific research community Writing a scientific bibliography Writing and presenting for research Designing and interpreting experimental work Ethics of research The student will also carry out a project (miniature internship) in the field of data science (various topics will be proposed by the faculty teaching in the master), with emphasis on problem formalisation and bibliography, and light experimentation.
Méthodes d'enseignement	
Langue d'enseignement	Anglais
Bibliographie	

	Semantic knowledge representation
Lieu d'enseignement	Polytech Nantes
Niveau	Master
Semestre	3
Responsable de l'UE	
Volume horaire total	**TOTAL :** 24h Répartition : **CM :** 0h **TD :** 0h **CI :** 14h **TP :** 7.5h **EAD :** 2.5h
Place de l'enseignement	
UE pré-requise(s)	
Parcours d'études comprenant l'UE	M2 Data Science (DS)
Evaluation	

Pondération pour chaque matière	Semantic knowledge representation **100%**
Obtention de l'UE	
Programme	
Objectifs (résultats d'apprentissage)	Upon completion, the student will be able to : - understand the different principle and concepts of knowledge representation : languages, annotation, reasoners. - to design a simple model of ontology with owl and protege. - to perform data annotation from an ontology with a triple based representation for data (rdf), and perform inference with a reasoner (hermit, sparql) - to apply these web 3.0 technology on knowledge bases and data in the web (linked open data)
Contenu	In the frame of web 3.0, semantic knowledge representation is concerned with logic modeling of knowledge with ontologies (vocabulary for concepts and properties), model instantiation on entities (data annotation), and performing logic computation and inference (reasonings) depending on a goal. The course introduce the principles and models and tools to model, annotate and make reasonings. These concepts are applied on linked data in order to process the data and knowledge stored into the web.
Méthodes d'enseignement	
Langue d'enseignement	Anglais
Bibliographie	

	Probabilistic Graphical Models and statistical relational learning
Lieu d'enseignement	
Niveau	Master
Semestre	3
Responsable de l'UE	
Volume horaire total	**TOTAL :** 24h Répartition : **CM :** 0h **TD :** 0h **CI :** 14h **TP :** 7.5h **EAD :** 2.5h
Place de l'enseignement	
UE pré-requise(s)	
Parcours d'études comprenant l'UE	M2 Data Science (DS)
Evaluation	
Pondération pour chaque matière	Probabilistic Graphical Models and statistical relational learning **100%**
Obtention de l'UE	
Programme	
Objectifs (résultats d'apprentissage)	Upon completion, the student will - be able to model simple problems with simple probabilistic graphical models such as Bayesian networks or Markov networks - understand probabilistic inference and parameter/Structure learning algorithm dedicated to such models - understand extensions of PGMs dealing with time problems or relational data
Contenu	Probabilistic graphical models (PGMs) are an interesting framework for encoding probability distributions over complex domains. These representations sit at the intersection of statistics and computer science, relying on concepts from probability theory, graph algorithms, machine learning, and more. This course describes two basic PGM representations: Bayesian Networks, which rely on a directed graph; and Markov networks, which use an undirected graph. One last part of the course is dedicated to various extensions of these models (dynamic Bayesian networks, probabilistic relational models, Markov logic networks).
Méthodes d'enseignement	

Langue d’enseignement	Français
Bibliographie	

	Classification, representation learning and dimensionality reduction
Lieu d’enseignement	Polytech Nantes
Niveau	Master
Semestre	3
Responsable de l’UE	
Volume horaire total	**TOTAL :** 24h Répartition : **CM :** 0h **TD :** 0h **CI :** 14h **TP :** 7.5h **EAD :** 2.5h
Place de l’enseignement	
UE pré-requise(s)	
Parcours d’études comprenant l’UE	M2 Data Science (DS)
Evaluation	
Pondération pour chaque matière	Classification, representation learning and dimensionality reduction **100%**
Obtention de l’UE	
Programme	
Objectifs (résultats d'apprentissage)	Upon completion, the student will - face the problem of human crafted features and observe the benefit of automatic feature learning - understand cutting edge representation learning algorithms applied in data science - know how to use representation learning methods that scale well on a variety of (un)labeled, (multi)-modal, relational and heterogeneous data. - be able to tackle a new data given by an application or a new problem described in a scientific paper, and apply the aforementionned methods on it
Contenu	Prerequisites : linear algebra, probability, statistics Not mandatory (they will be partly covered during the first courses) but valuable : basic machine learning (classification, regression), convex optimization 1. Intro [Motivation, definitions, terminology, review linear algebra, probability and optimization, regression] 2. Subspace learning [principal component analysis (PCA), statistical and geometrical viewpoint, indep. component analysis (ICA)] 4. Manifold learning [MDS, ISOMAP, t-sne and other unsupervised manifold methods] 5. Deep learning [restricted Boltzmann machines, auto encoders, deep belief networks, convolutional neural networks, recurrrent neural networks] 6. Metric learning [(non)-linear, global/local, constraints setting, structured data] Project : Students should form groups of 2-4 members. A list of candidate papers will be posted, and each group should pick one from the list. Each group is required to give an oral presentation about the content of the paper in the last two weeks, and submit a report at the end. The report should include at the minimum a summary of the method/framework, and experimental results obtained by playing the code published along with the paper. Division of work should be determined by the members. Grading will be based on a project assignment (50%) and a final exam (50%).
Méthodes d’enseignement	
Langue d’enseignement	Anglais
Bibliographie	

	Cluster analysis and indexing
Lieu d’enseignement	Polytech Nantes
Niveau	Master
Semestre	3

Responsable de l’UE	GELGON MARC
Volume horaire total	**TOTAL :** 24h Répartition : **CM :** 0h **TD :** 0h **CI :** 14h **TP :** 7.5h **EAD :** 2.5h
Place de l’enseignement	
UE pré-requise(s)	
Parcours d’études comprenant l’UE	M2 Data Science (DS)
Evaluation	
Pondération pour chaque matière	Cluster analysis and indexing **100%**
Obtention de l’UE	Note : le public attendu pour ce parcours "Data Science" est mixte : étudiants ayant suivi le M1 du master informatique de l'université de Nantes, étudiants venant de l'étranger (cible spécifique explicite pour ces parcours internationaux). Il est prévu de personnaliser, dans la promotion, les modalités pédagogiques et contenus enseignés pour tenir compte de cette hétérogénéité. Cette remarque vaut pour la plupart des UE de ce parcours.
Programme	
Objectifs (résultats d'apprentissage)	Upon completion, the student will be able to : - recognize an unsupervised or semi-supervised classification problem from real-world problems and characterize it - evaluate the suitabilty of the main mathematical models and techniques to address this task - design, carry out and criticize experiments on real data sets - develop software using relevant libraries for carrying out new classification problems and experiments - start exploiting scientific resources such as scientific papers
Contenu	Advanced clustering, co-clustering, semi-supervised classification. Probabilistic mixture models, topic models Matrix factorization Typical applications and experiments.
Méthodes d’enseignement	
Langue d’enseignement	Anglais
Bibliographie	

	Text and sequential pattern mining
Lieu d’enseignement	Polytech Nantes
Niveau	Master
Semestre	3
Responsable de l’UE	
Volume horaire total	**TOTAL :** 24h Répartition : **CM :** 0h **TD :** 0h **CI :** 14h **TP :** 7.5h **EAD :** 2.5h
Place de l’enseignement	
UE pré-requise(s)	
Parcours d’études comprenant l’UE	M2 Data Science (DS)
Evaluation	
Pondération pour chaque matière	Text and sequential pattern mining **100%**
Obtention de l’UE	
Programme	

Objectifs (résultats d'apprentissage)	Students will be able to: - run through the main algorithms for mining sequential patterns, episodes and processes from a small dataset. - choose interestingness measures appropriate to the data and to the analyst' goals. - preprocess text corpuses and characterize them with patterns.
Contenu	This teaching unit explores techniques for discovering and assessing patterns and structures from sequential data (temporal event sequences, texts, biological sequences, etc.). It focuses on applications for text mining and process mining. Outline: Sequential pattern mining - episode mining - sequential pattern mining - constraint-based mining - pattern assessment Text Mining - preprocessing methods - similarities - emerging patterns Process Mining - Process model - Process discovery - Conformance checking
Méthodes d’enseignement	
Langue d’enseignement	Anglais
Bibliographie	

	Pattern mining and social network analysis
Lieu d’enseignement	Polytech Nantes
Niveau	Master
Semestre	3
Responsable de l’UE	
Volume horaire total	**TOTAL :** 24h Répartition : **CM :** 0h **TD :** 0h **CI :** 14h **TP :** 7.5h **EAD :** 2.5h
Place de l’enseignement	
UE pré-requise(s)	
Parcours d’études comprenant l’UE	M2 Data Science (DS)
Evaluation	
Pondération pour chaque matière	Pattern mining and social network analysis **100%**
Obtention de l’UE	
Programme	
Objectifs (résultats d'apprentissage)	
Contenu	
Méthodes d’enseignement	
Langue d’enseignement	Anglais
Bibliographie	

	Visual analytics

Lieu d’enseignement	Polytech Nantes
Niveau	Master
Semestre	3
Responsable de l’UE	
Volume horaire total	**TOTAL :** 24h Répartition : **CM :** 0h **TD :** 0h **CI :** 14h **TP :** 7.5h **EAD :** 2.5h
Place de l’enseignement	
UE pré-requise(s)	
Parcours d’études comprenant l’UE	M2 Data Science (DS)
Evaluation	
Pondération pour chaque matière	Visual analytics **100%**
Obtention de l’UE	
Programme	
Objectifs (résultats d'apprentissage)	At the end of the course the student will be able to - choose an adequate visual representation for a dataset, with regards to the human perception and cognitive systems and the insights that are needed - design an interactive visual analytics system, combining analytical techniques and data representation for the various tasks a domain analyst needs to carry - carry out the evaluation of an interactive visualisation system, choosing from various evaluation means - engage in discussions on the latest theoretical research topics and challenges
Contenu	1- Introduction to data visualization human factors, marks and visual channels, mappings, common errors, classical data visualisations, tools, etc. 2- Data representation techniques clustering and dimensionality reduction, trees and network representations, time series representations, 3D representations, etc. 3- Designing and evaluating interactive visual analytics systems analytics loop, types of interactions in visual analytics, design methods, evaluation methods, etc. 4- Projects 10 courses of 2h each. Each student will present one or two times during the course. The project will consist in choosing a data analysis algorithm / method, and design (paper + Tableau) an interactive dataviz for an analyst to study the data, the results of the algorithm as well as the quality metrics, while allowing to relaunch the algorithm. Written examination at the end.
Méthodes d’enseignement	
Langue d’enseignement	Anglais
Bibliographie	

	Data dependencies and data integration
Lieu d’enseignement	Polytech Nantes
Niveau	Master
Semestre	3
Responsable de l’UE	
Volume horaire total	**TOTAL :** 21.5h Répartition : **CM :** 0h **TD :** 0h **CI :** 14h **TP :** 7.5h **EAD :** 0h
Place de l’enseignement	
UE pré-requise(s)	

Parcours d’études comprenant l’UE	M2 Data Science (DS)
Evaluation	
Pondération pour chaque matière	Data dependencies and data integration **100%**
Obtention de l’UE	
Programme	
Objectifs (résultats d'apprentissage)	
Contenu	- Part I: * A Short Review of the Relational Data Model: SQL, RA, RC, CQ, FO * Functional Dependencies and Inclusion Dependencies: Armstrong's Axioms, the Implication Problem * Database Design: BCNF, 3NF, Decomposition, Chase test - Part II * FD discovery: TANE, FD_Mine, Dep-Miner, CORDS, FastFDs * Extension to Approximate FD's discovery * Conditional FD's and 33 other Relaxations! - Part III * Data Integration: egds, tgds, G/L-AV * Schema Mapping: GLAV * Data Exchange: universal instance, certain answers * Query-Answering Using Views
Méthodes d’enseignement	
Langue d’enseignement	Anglais
Bibliographie	

	Data economics, law and ethics
Lieu d’enseignement	Polytech Nantes
Niveau	Master
Semestre	3
Responsable de l’UE	
Volume horaire total	**TOTAL :** 24h Répartition : **CM :** 0h **TD :** 0h **CI :** 21.5h **TP :** 0h **EAD :** 2.5h
Place de l’enseignement	
UE pré-requise(s)	
Parcours d’études comprenant l’UE	M2 Data Science (DS)
Evaluation	
Pondération pour chaque matière	Data economics, law and ethics **100%**
Obtention de l’UE	
Programme	
Objectifs (résultats d'apprentissage)	Upon completion, students will be able to : - relate data analysis tasks and economic motivations and models in the digital economy - relate the legal and privacy and algorithmics of privacy-preserving and discrimination-preserving technologies - take advantage of mathematical models and software tools that exploit personal data to create economic value (recommendation, crowdsourcing)
Contenu	Economics of data Open data Models and techniques for recommender systems Models and techniques for crowdsourcing Anonymization techniques for privacy-preserving data publishing Detecting and preventing discrimination

Méthodes d’enseignement	
Langue d’enseignement	Anglais
Bibliographie	

	Conferences and invited courses (DS)
Lieu d’enseignement	Polytech Nantes
Niveau	Master
Semestre	3
Responsable de l’UE	
Volume horaire total	**TOTAL :** 24h Répartition : **CM :** 0h **TD :** 0h **CI :** 21.5h **TP :** 0h **EAD :** 2.5h
Place de l’enseignement	
UE pré-requise(s)	
Parcours d’études comprenant l’UE	M2 Data Science (DS)
Evaluation	
Pondération pour chaque matière	Conferences and invited courses (DS) **100%**
Obtention de l’UE	Les étudiants doient obligatoirement assister aux conférences pour valider l'UE.
Programme	
Objectifs (résultats d'apprentissage)	This series of presentations and discussions will open their mind of students to new topics, applications and speakers, and stimulate them for chosing their way into the field of data science.
Contenu	This course is defined with regular short conferences with researcher or industrialist. The program is defined every year in accordance with the topical issues.
Méthodes d’enseignement	
Langue d’enseignement	Anglais
Bibliographie	

	French langage and European culture
Lieu d’enseignement	
Niveau	Master
Semestre	3
Responsable de l’UE	
Volume horaire total	**TOTAL :** 48h Répartition : **CM :** 0h **TD :** 48h **CI :** 0h **TP :** 0h **EAD :** 0h
Place de l’enseignement	
UE pré-requise(s)	
Parcours d’études comprenant l’UE	M2 Visual Computing (VICO),M2 Data Science (DS)
Evaluation	
Pondération pour chaque matière	French langage and European culture **100%**

Obtention de l'UE	
Programme	
Objectifs (résultats d'apprentissage)	Upon completing the course in French as a foreign language, students will be able to : • introduce themselves giving basic information about their country, family, studies • greet native speakers in an appropriate way • interact with native speakers in routine basic tasks (shopping, ordering at a restaurant) • fill in registration forms
Contenu	Grammar • tenses : présent, futur proche, passé proche, the imperative • questions and negations • pronouns (subject and object) • some irregular verbs (être, avoir, prendre, faire, aller, venir) Vocabulary • days of the week, months, numbers, jobs, food items, clothes, modes of transportation, nationalities... Phonetics • difficult vowel and consonant sounds • liaisons Culture and civilization • the outdoor market, Christmas, food and meals, the pace of life in France
Méthodes d'enseignement	
Langue d'enseignement	Français
Bibliographie	

	Internship (DS)
Lieu d'enseignement	Polytech Nantes
Niveau	Master
Semestre	4
Responsable de l'UE	
Volume horaire total	**TOTAL :** 0h Répartition : **CM :** 0h **TD :** 0h **CI :** 0h **TP :** 0h **EAD :** 0h
Place de l'enseignement	
UE pré-requise(s)	
Parcours d'études comprenant l'UE	M2 Data Science (DS)
Evaluation	
Pondération pour chaque matière	Internship (DS) **100%**
Obtention de l'UE	
Programme	
Objectifs (résultats d'apprentissage)	At the end of the intership, the student should be able to: 1. 1. Identify and refine a research question or puzzle within an existing field of scientific inquiry and devise a plan for investigating it. 2. Formulate a program of reading in consultation with a professional scientist to provide context for the investigation 3. Develop a time-line for the research project and manage work to that time-line 4. Communicate research results –both orally and in writing – in a style consistent with scientific standards 5. Work as part of a research team
Contenu	During the internship, the student will conduct a research work either in a university lab or in the R&D department of a private company.
Méthodes d'enseignement	
Langue d'enseignement	Anglais

Bibliographie	

Dernière modification par ISABELLE BEAUDET, le 2020-05-29 17:41:42