

Semestre 3

N°	Unité d'enseignement (UE)	Nature et code de l'UE		Elément constitutif d'UE (ECUE)	Volume horaire			Nombre de Crédits accordés		Coefficients		Modalité d'évaluation	
					C	TD	TP	ECUE	UE	ECUE	UE	Contrôle Continu	Mixte
11	Deep Learning	Fondamentale M1MA/DS11	MDS111	Machine Learning 2	28	14		4	7	2	3.5		X Examen 1.5
				MDS112	Optimisation pour l'apprentissage	28	14		3		1.5		
12	Intelligence artificielle	Fondamentale M1MA/DS12	MDS121	I.A. on the cloud	28	14		3	7	1.5	3.5		X Examen 1.5
				MDS122	Recherche Monte Carlo et logique de description	28	14		4		2		
13	Processus stochastique 2 et modélisation des données	Fondamentale M1MA/DS13	MDS131	Processus de Poisson et Fiabilité des logiciels	28	14		4	7	2	3.5		X Examen 1.5
				MDS132	Flux et fouilles de données	28	14		3		1.5		
14	Sciences des données et Big Data	Optionnelle M1MA/DS14	MDS141	Data Wrangling	10.5	10.5		3	5	1.5	2.5		X Examen 1.5
				MDS142	Data visu et projet sciences des données	10.5	10.5		2			1	
15	Programmation avancée avec Python	Transversale M1MA/DS15	MDS151	Programmation avancée avec Python	14	14			4		2	X	
Total : 322h /semestre					203h	91h	28h		30		15		

Semestre 4

N°	Unité d'enseignement (UE)	Type de l'UE (Obligatoire / Optionnelle)	Elément constitutif d'UE (ECUE)	Volume des heures de formation présentielle (14 semaines)				Nombre de Crédits accordés		Coefficients		Modalité d'évaluation	
				Cours	TD	TP	Autres	ECUE	UE	ECUE	UE	CC	Mixte
1	Activité pratique	Obligatoire	M. de Recherche : Mémoire de Recherche					30	30	15	15		
Total								30	30	15	15		

Cours Master de recherche en Mathématiques et Applications : Parcours Data Science

Semestre 3

- Deep Learning
- Intelligence artificielle
- Processus stochastique 2 et modélisation des données
- Sciences des données et Big Data
- Programmation avancée avec Python

Semestre 4

- Mémoire de recherche

Contenu des programmes des modules en 2^{ème} année

Programme de l'UE : Deep Learning

Machine Learning 2 :

- Code : MDS111
- Cours et TD : 42H
- Crédits : 4
- Coefficients : 2

L'objectif de ce cours est de fournir aux étudiants les concepts et outils fondamentaux pour développer et analyser des algorithmes d'apprentissage automatique. Le cours présentera les fondements théoriques de l'apprentissage automatique, passera en revue les algorithmes les plus performants avec leurs garanties théoriques et discutera de leur application dans des problèmes du monde réel.

Les thèmes abordés sont :

1. Introduction aux différents paradigmes du ML et des applications
2. Introduction aux différents paradigmes du ML et des applications
3. Théorie de l'apprentissage informatique : Modèle PAC, VC-dimension, complexité Rademacher,...
4. Enseignement supervisé : Régression logistique et au-delà
5. Apprentissage non supervisé : Réduction de la dimensionnalité : PCA, ICA, Projections aléatoires, Kernel PCA, ISOMAP, LLE, Estimation de la densité, SE, Regroupement spectral
6. Apprentissage en ligne
7. Algorithmes multiclassés et de classement
8. Séances pratiques- Complexité Rademacher,...
9. Enseignement supervisé
10. Régression logistique et au-delà

Optimisation pour l'apprentissage :

- Code : MDS112
- Cours et TD : 42H
- Crédits : 3
- Coefficients : 1.5

L'optimisation est au cœur des avancées les plus récentes en apprentissage automatique. Cela inclut bien sûr les méthodes les plus basiques (régression linéaire, SVM et méthodes du noyau). C'est également la clé de l'explosion récente de l'apprentissage en profondeur qui sont des approches de pointe pour résoudre des problèmes supervisés et non supervisés en imagerie, vision et traitement automatique des langages. Ce cours passera en revue les fondements mathématiques, les

méthodes algorithmiques sous-jacentes et présentera quelques applications modernes d'un large éventail de techniques d'optimisation.

Le cours sera composé à la fois de cours magistraux classiques et de sessions numériques en Python. La première partie couvre les méthodes de base de l'optimisation lisse (descente de gradient) et optimisation convexe (condition d'optimalité, optimisation contrainte, dualité). La deuxième partie présentera des méthodes plus avancées (optimisation non lisse, optimisation SDP, points intérieurs et méthodes proximales). La dernière partie couvrira les méthodes à grande échelle (descente de gradient stochastique), la différenciation automatique (en utilisant le framework python moderne) et leur application aux réseaux de neurones (réseaux peu profonds et profonds).

Programme de l'UE : Intelligence artificielle

I.A. on the cloud :

- **Code : MDS121**
- **Cours et TD : 42H**
- **Crédits : 3**
- **Coefficients : 1.5**

L'objectif principal de ce cours est de présenter aux étudiants et de leur donner la possibilité d'acquérir des connaissances sur les architectures Cloud typiques pour prendre en charge toutes les phases du traitement de données IA typique : stockage et préparation des données ainsi que déploiement et exécution d'algorithmes d'apprentissage automatique. Une attention particulière sera accordée aux architectures cloud typiques et à la manière dont elles peuvent assurer un traitement optimal des données dans les pipelines IA, en prenant en compte le coût monétaire des ressources parmi d'autres paramètres traditionnels.

Recherche Monte Carlo et logique de description :

- **Code : MDS122**
- **Cours, TD et TP : 42H**
- **Crédits : 4**
- **Coefficients : 2**

Contenu de l'enseignement.

Ce cours est une introduction aux méthodes dites de Monte-Carlo. Ces méthodes sont utilisées pour calculer des espérances, et par extension, des intégrales par simulation.

L'objectif de ce cours est non seulement de fournir les bases théoriques des méthodes de Monte-Carlo, mais aussi de fournir les outils permettant leur utilisation pratique à travers des TP.

Points abordés :

Le cours couvre les sujets suivants :

1. Introduction de la méthode de Monte-Carlo
2. Techniques de réduction de variance
3. Introduction aux suites à discrétion faible

Programme de l'UE : Processus stochastique 2 et modélisation des données

Processus de Poisson et fiabilité des logiciels :

- **Code : MDS131**
- **Cours et TP : 42H**
- **Crédits : 4**
- **Coefficients : 2**

Compétences à acquérir

Introduction des processus à temps continus fondamentaux en probabilités, tels que les chaînes de Markov à espace d'états dénombrable et les processus de renouvellement. Présentation de la théorie du renouvellement et de la théorie des files d'attente.

Description du contenu de l'enseignement

Chapitre 1 : Processus de comptage

1. Définitions et propriétés classiques.
2. Processus de Poisson.
3. Propriétés caractéristiques d'un processus de Poisson : PAIS....
4. Processus de Poisson composée.

Chapitre 2 : Introduction à la théorie de renouvellement

1. Processus de renouvellement.
2. Produit de convolution des mesures.
3. Equation de renouvellement.
4. Théorème de renouvellement clé et théorème de Blackwell's.
5. Cas exponentiel.

Chapitre 3 : Processus markoviens de sauts

1. Définition et propriétés classiques
2. Construction'
3. Semi-groupe et équations de Kolmogorov.
4. Récurrence et transience d'un processus markovien de sauts
5. Mesure invariante et propriétés classiques.

Chapitre 4 : Processus de naissance et mort : application aux files d'attentes

1. Définition et propriétés élémentaires.
2. Notion des files d'attente et exemples : M/M/1; M/M/s; M/M/oo; M/M/1/k.
3. Une file un peu plus compliquée.
4. Un modèle simple de population

Flux et fouilles de données :

- **Code : MDS132**
- **Cours, TD et TP : 42H**
- **Crédits : 3**
- **Coefficients : 1.5**

Ce cours a pour objectif de décrire les démarches, méthodes et outils utilisés pour représenter les données complexes et multiples issues des grandes masses de données en visuels simples à comprendre et à interpréter, notamment pour les utilisateurs métier et pour les décideurs.

Le cas de l'apport de la visualisation des données sous différentes formes graphiques lors de la préparation des données en amont de la modélisation et de l'utilisation des modèles et algorithmes de Machine Learning sera également développé.

Le cours s'appuie sur de nombreux exemples puisés dans les domaines de la finance, de la santé, du marketing et des travaux pratiques sur des cas concrets sont également prévus.

Programme de l'UE : Sciences des données et Big Data

Data Wrangling :

- **Code : MDS141**
- **Cours et TD : 21H**
- **Crédits : 3**
- **Coefficients : 1,5**

Objectifs

Ce cours a pour objectif d'enseigner une méthodologie pour diagnostiquer et corriger les problèmes dus à la non-qualité des données, mettre en œuvre une démarche qualité des données et mesurer ses effets. Il donne également un aperçu des outils existants et de leur utilisation.

Contenu de l'enseignement

1. Les différentes sources de données et leur exploitation.
2. Mesure de la qualité des données et principales méthodes existantes.
3. Coût de la qualité.
4. Méthodes d'identification et de correction des données suivant leur type (manquantes, aberrantes, erronées, ...).
5. Indicateurs et suivi qualité des données.
6. Amélioration de la qualité des données.
7. Les outils logiciels et la qualité des données

Data visu et projet science des données :

- **Code : MDS142**
- **Cours et TD : 21H**
- **Crédits : 2**
- **Coefficients : 1**

Ce cours a pour objectif de décrire les démarches, méthodes et outils utilisés pour représenter les données complexes et multiples issues des grandes masses de données en visuels simples à comprendre et à interpréter, notamment pour les utilisateurs métier et pour les décideurs.

Le cas de l'apport de la visualisation des données sous différentes formes graphiques lors de la préparation des données en amont de la modélisation et de l'utilisation des modèles et algorithmes de Machine Learning sera également développé.

Le cours s'appuie sur de nombreux exemples puisés dans les domaines de la finance, de la santé, du marketing et des travaux pratiques sur des cas concrets sont également prévus.

Les séances de TD sur machines seront consacrées à un training sur des logiciels de visualisation des données comme Tableau/ Power BI.

Projet sur les données

L'objectif de ce module est de fournir aux étudiants une expérience pratique sur un nouveau défi de science des données/IA en utilisant les outils et techniques de l'état de l'art abordés lors des autres cours de ce master.

Les étudiants inscrits dans cette classe formeront des groupes et choisiront un sujet parmi une liste de sujets proposés dans les domaines de base du master tels que l'apprentissage supervisé ou non supervisé, la recommandation, l'IA de jeu, la science des données distribuée ou parallèle, etc. Les sujets consisteront généralement à appliquer une technique bien établie sur un nouveau défi de science des données ou à appliquer des résultats de recherche récents sur un défi classique de science des données.

Quoi qu'il en soit, chaque sujet viendra avec son propre nouveau défi scientifique à relever. À la fin du module, les étudiants feront une présentation orale pour démontrer leur méthodologie et leurs résultats. Une forte rigueur scientifique ainsi que de très bonnes compétences en ingénierie et en communication seront nécessaires pour réussir ce module.

Programme de l'UE : Programmation avancée avec Python

- **Code : MDS151**
- **Cours et TD: 28h**
- **Crédits : 4**
- **Coefficients : 2**

Le but de cette UE est de résoudre des problèmes scientifiques avec python afin de découvrir la démarche de la modélisation mathématique. Les étudiants devront effectuer des mini projets sous forme de comptes rendus. L'objectif de chaque projet est de formuler et analyser mathématiquement le problème mathématique posé, et de proposer et mettre

en œuvre des algorithmes afin de modéliser et résoudre ce problème. Cette UE vise à développer l'autonomie des étudiants par le travail en groupe sur des projets ouverts et des questions de recherche.

Evaluation:

Chaque étudiant doit préparer et présenter un mini projet qui sera comptabilisé sous forme d'examen

1. Rappels sur la programmation avec python

2. 2Présentation et répartition des projets

3. THEMATIQUES

3.1Analyse exploratoire des données

3.2Modélisation, mise en œuvre des algorithmes de machine Learning sur des données réelles : Analyse factorielle, SVM, régression linéaire multiple, régression généralisée, gradient boosting, régression lasso

3.3Problèmes de classification : Kmeans, ACP, KNN...

3.4Problèmes d'optimisation

3.5EDP avec python