

**REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET  
POPULAIRE**

**MINISTRE DE L'ENSEIGNEMENT SUPERIEUR  
ET DE LA RECHERCHE SCIENTIFIQUE**

# **Canevas**

## **OFFRE DE FORMATION MASTER**

### **ACADÉMIQUE**

**2020 - 2021**

<b>Etablissement</b>	<b>Faculté</b>	<b>Département</b>
<b>Université Alger1</b>	<b>Sciences</b>	<b>Mathématiques et Informatique</b>

<b>Domaine</b>	<b>Filière</b>	<b>Spécialité</b>
<b>Mathématiques et Informatique</b>	<b>Informatique</b>	<b>Analyse et Sciences de Données</b>

# SOMMAIRE

I – Fiche d’identité du Master	3
1 - Localisation de la formation :	5
2 - Partenaires extérieurs	4
3 – Contexte et objectifs de la formation	4
A - Conditions d’accès	4
B - Objectifs de la formation	4
C - Profils et compétences visées	5
D – Potentialités régionales et nationales d'employabilité	5
E – Grandes classes d'applications industrielles du Master	5
F – Passerelles vers les autres spécialités	6
G – Indicateurs de suivi de la formation	6
H – Moyens humains disponibles	6
4 – Moyens humains disponibles	7
A : Equipe pédagogique interne mobilisée pour la spécialité :	7
B : Equipe pédagogique externe mobilisée pour la spécialité :	9
C : Synthèse globale des ressources humaines mobilisées pour la spécialité DSA :	11
5 – Moyens matériels spécifiques à la spécialité	12
A - Laboratoires Pédagogiques et Equipements : Fiche des équipements pédagogiques existants pour les TP de la formation envisagée (une fiche par laboratoire)	12
B - Terrains de stage et formations en entreprise (voir rubrique accords / conventions):	12
C - Laboratoire(s) de recherche de soutien au master :	13
D - Projet(s) de recherche de soutien au master :	13
E - Espaces de travaux personnels et TIC disponibles au niveau du département et de la faculté :	14
II – Fiche d’organisation semestrielle des enseignements	15
1 - Semestre 1	16
2 - Semestre 2	17
3 - Semestre 3	18
4 - Semestre 4	19
5- Récapitulatif global de la formation :	19
III - Programme détaillé par matière	20
IV - Avis et Visas des organes administratifs et consultatifs	45
V - Avis et Visa de la Conférence Régionale	45
VI - Avis et Visa du Comité Pédagogique National de Domaine (CPND)	46
VI - Avis et Visa du Comité Pédagogique National de Domaine (CPND)	46

## I – Fiche d'identité du Master

## **1 - Localisation de la formation :**

**Etablissement :** Université Alger1

**Faculté :** Sciences

**Département :** Mathématiques et Informatique

## **2 - Partenaires extérieurs**

- Autres établissements universitaires
- Université des sciences et de la technologie Houari Boumediene-USTHB
- Ecole Supérieure d'Informatique, ESI

## **3 – Contexte et objectifs de la formation**

### **A - Conditions d'accès**

Les étudiants ayant suivi une formation de Licence en « Mathématiques et Informatique » à parcours « Informatique », « Mathématiques Appliquées » ou une formation jugée équivalente peuvent candidater pour ce master. L'admission se fait sur dossier et en fonction des places disponibles. La candidature est ouverte pour tout étudiant détenteur de l'un des diplômes suivants :

- Licence en Systèmes Informatiques.
- Licence en Ingénierie des Système d'Information et du Logiciel
- Licence en Mathématiques Appliquées
- Tout titre reconnu équivalent.

### **B - Objectifs de la formation**

Ce Master offre aux étudiants des bases théoriques et pratiques solides dans le domaine des sciences de données. Ces bases leur permettront de s'insérer facilement dans le monde socio-économique ou d'intégrer des organismes de la recherche publique ou privée et de s'orienter vers le monde de la recherche théorique ou appliquée de haut niveau. En effet, depuis la fin des années 90 les scientifiques ont marqué le passage de l'ère des sciences informatiques vers l'ère des sciences des données. Suite à l'explosion exponentielle en termes de volume de données que connaît le monde, la majorité écrasante des grandes et moyennes entreprises ainsi que les startups possède un département ou une équipe travaillant dans le domaine de la data science.

Ce master propose une formation approfondie en sciences des données et leurs applications. Les diplômés peuvent assurer des tâches d'ingénieurs informaticiens en général et plus particulièrement les tâches de data analyst, data scientist et data engineer dans les établissements publics ou privés. Ils peuvent aussi candidater à un doctorat en Informatique, particulièrement en Sciences des données, en Intelligence Artificielle, en traitement automatiques des langues...

### **C - Profils et compétences visées**

- Acquisition de connaissances approfondies et solides en Informatique (Intelligence Artificielle, Traitement et Analyse de masses de données - Big data, Extraction de connaissances à partir de données complexes, représentation et visualisation de l'information complexe, etc.).
- Concevoir et développer des systèmes pour la gestion des données massives et hétérogènes.
- Maîtriser la conception et la mise en œuvre de systèmes d'apprentissage à partir des données brutes jusqu'à l'évaluation.
- Veille technologique et suivi de l'évolution de l'informatique, avec la présence de ces grandes masses de données, offre la possibilité de repenser le paradigme informatique.
- Le diplômé doit savoir appréhender des interfaces différentes liées à la mobilité et à des environnements hétérogènes. Les transferts d'information deviennent complexes et nécessitent des compétences mêlant le réseau et la sécurité. L'extraction d'information requiert des techniques spécifiques et des algorithmes évolués. Quant à l'analyse de grandes masses de données, elle ne peut se faire sans connaissance de modèles et techniques mathématiques et informatiques, tant pour la pertinence de l'information résultante que pour sa représentation et sa visualisation.
- Initiation à la recherche scientifique : démarche de recherche, capacité d'abstraction, autonomie, esprit de synthèse, ...
- Apprentissage de la rédaction et présentation informatique avec une démarche scientifique et maîtrise des outils informatiques avancés.
- Capacité d'analyse du besoin du donneur d'ordre pour présenter le tableau de bord d'aide à la décision le plus pertinent possible.

### **D – Potentialités régionales et nationales d'employabilité**

A l'issue de la formation, les étudiants peuvent aussi bien travailler en tant que diplômé en master Data Science ou poursuivre leurs études en thèse de doctorat en informatique. Le manque d'ingénieurs spécialisés sur tout le territoire algérien, le continent africain et dans le monde, offre de larges possibilités d'employabilités aux diplômés de ce Master en Informatique.

Au terme du Master les métiers visés sont notamment :

- Data scientist, Data engineer ou Data analyst
- Ingénieur en Intelligence Artificielle
- Administrateur d'entrepôts et de bases de données complexes
- Gestionnaire d'applications et/ou de données massives
- Architecte de données
- Concepteur/développeur d'applications Big Data
- Ingénieur en recherche et développement

### **E – Grandes classes d'applications industrielles du Master**

- Santé
- Énergie
- Recherche sur Internet

- Publicité ciblée
- Astrophysique et météorologie
- Reconnaissance avancée de formes
- Reconnaissance de la parole
- Planification en transport
- Gaming
- Réalité virtuelle et augmentée
- Robotique

#### **F – Passerelles vers les autres spécialités**

Possibilité de passage vers d'autres spécialités d'informatique :

- Bases de données.
- Systèmes d'informations.

#### **G – Indicateurs de suivi de la formation**

- Épreuves de courte durée (Contrôle continu en cours de semestre).
- Épreuves finales à la fin de chaque semestre.
- Mémoires et soutenances.
- Taux de réussite en M1 et M2.
- Nombre d'étudiants inscrits en Doctorat.
- Nombre d'étudiants recrutés à l'issue de la formation.

#### **H – Moyens humains disponibles**

**Capacité d'encadrement** : 30 étudiants en M1. Ceci dépendra par la suite de l'évolution de l'encadrement en termes de nombre et de spécialités des enseignants en activité au sein du département, mais aussi de leur disponibilité.

**C : Synthèse globale des ressources humaines mobilisées pour la spécialité DSA :**

<b>Grade</b>	<b>Effectif Interne</b>	<b>Effectif Externe</b>	<b>Total</b>
<b>Professeurs</b>	2	1	<b>3</b>
<b>Maîtres de Conférences (A)</b>	1	0	<b>1</b>
<b>Maîtres de Conférences (B)</b>	12	5	<b>17</b>
<b>Maître Assistant (A)</b>	8	1	<b>9</b>
<b>Maître Assistant (B)</b>	0	0	<b>0</b>
<b>Ingénieurs d'informatique</b>	5	0	<b>5</b>
<b>Total</b>	<b>28</b>	<b>7</b>	<b>35</b>

## 5 – Moyens matériels spécifiques à la spécialité

### A - Laboratoires Pédagogiques et Equipements : Fiche des équipements pédagogiques existants pour les TP de la formation envisagée (une fiche par laboratoire)

**Intitulé du laboratoire :**

**Salles de TP :**

**Capacité en étudiants :**

N°	Intitulé de l'équipement	Nombre	Observations
01	Ordinateurs de Bureau	32	15, 32G de RAM, Carte Graphique NVIDIA GTR 2080
02	Serveur central	01	10eme Gén
03	Accès Internet	32	
04	Data Show	05	

### B - Terrains de stage et formations en entreprise (voir rubrique accords / conventions):

Lieu du stage	Nombre d'étudiants	Durée du stage
Centre de développement des Technologies Avancées (CDTA)	6	06 mois
Centre de développement des Technologies Avancées (CERIST)	6	06 mois
Laboratoire FRATER RAZES	12	06 mois
BNA	6	06 mois



**E - Espaces de travaux personnels et TIC disponibles au niveau du département et de la faculté :**

L'université d'Alger 1 dispose de :

- Une salle de lecture de 600 places à la bibliothèque de l'université.
- Une deuxième salle de lecture de 100 places à la faculté des sciences.
- Une grande salle de travail au département Mathématiques et Informatique équipée de 30 ordinateurs connectés au réseau de l'université pour l'usage des étudiants.

## **II – Fiche d'organisation semestrielle des enseignements**

# 1 - Semestre 1

Unité d'Enseignement	VHS	V.H hebdomadaire				Coeff	Crédits	Mode d'évaluation	
	14-sem	C	TD	TP	Travail personnel			Contrôle continu	Examen
<b>UE fondamentales</b>									
<b>UEF1.1</b>	<b>147h</b>	<b>4h30</b>	<b>3h</b>	<b>3h</b>	<b>9h</b>	<b>8</b>	<b>13</b>		
UEF1.1.1 : Probabilités		1h30	1h30	1h30	3h	3	5	40 %	60 %
UEF1.1.2 : Mathématiques pour l'ingénierie		1h30	1h30		3h	2	4	40 %	60 %
UEF1.1.3 : Statistique exploratoire		1h30		1h30	3h	3	4	40 %	60 %
<b>UEF1.2</b>	<b>168h</b>	<b>4h30</b>	<b>1h30</b>	<b>6h</b>	<b>9h</b>	<b>8</b>	<b>14</b>		
UEF1.2.1 : Apprentissage et Complexité		1h30	1h30	1h30	3h	3	5	40 %	60 %
UEF1.2.2 : Machine Learning		1h30		3h00	3h	3	5	40 %	60 %
UEF1.2.3 : Data warehouse et OLAP		1h30		1h30	3h	2	4	40 %	60 %
<b>UE méthodologie</b>									
<b>UEM1.1</b>	<b>42h</b>	<b>1h30</b>	<b>1h30</b>	<b>0h</b>	<b>3h</b>	<b>1</b>	<b>2</b>		
UEM1.1.1: Marketing, techniques de communication, aspects juridiques		1h30	1h30		3h	1	2	40 %	60 %
<b>UE transversale</b>									
<b>UET1.1</b>	<b>21h</b>	<b>0h</b>	<b>0h</b>	<b>1h30</b>	<b>1h30</b>	<b>1</b>	<b>1</b>		
UET1.1.1 : Anglais pour la communication professionnelle				1h30	1h30	1	1		100 %
<b>Total Semestre 1</b>	<b>378h</b>	<b>10h30</b>	<b>6h</b>	<b>10h30</b>	<b>22h30</b>	<b>18</b>	<b>30</b>		

## 2 - Semestre 2

Unité d'Enseignement	VHS	V.H hebdomadaire				Coeff	Crédits	Mode d'évaluation	
	14- sem	C	TD	TP	Travail personnel			Contrôle continu	Examen
<b>UE fondamentales</b>									
<b>UEF2.1</b>	<b>147h</b>	<b>4h30</b>	<b>3h</b>	<b>3h</b>	<b>9h</b>	<b>7</b>	<b>13</b>		
<b>UEF2.1.1:</b> Statistique Inférentielle		1h30	1h30		3h	2	4	40 %	60 %
<b>UEF2.1.2:</b> Calcul Intensif		1h30		1h30	3h	2	4	40 %	60%
<b>UEF2.1.3:</b> Optimisation		1h30	1h30	1h30	3h	3	5	40 %	60 %
<b>UEF2.2</b>	<b>84h</b>	<b>3h</b>	<b>0h</b>	<b>3h</b>	<b>6h</b>	<b>6</b>	<b>9</b>		
<b>UEF2.2.1:</b> Data Mining		1h30		1h30	3h	3	5	40 %	60 %
<b>UEF2.2.2:</b> Extraction d'Information dans les Documents textes, audio, vidéo		1h30		1h30	3h	3	4	40 %	60%
<b>UE méthodologies</b>									
<b>UEM2.1</b>	<b>84h</b>	<b>3h</b>	<b>0h</b>	<b>3h</b>	<b>6h</b>	<b>4</b>	<b>7</b>		
<b>UEM2.1.1:</b> Principes et Méthodes des systèmes de Bases de Données Réparties		1h30		1h30	3h	2	3	40 %	60 %
<b>UEM2.1.2:</b> Simulation pour les systèmes complexes		1h30		1h30	3h	2	4	40 %	60 %
<b>UE transversale</b>									
<b>UET2.1</b>	<b>21h</b>	<b>0h</b>	<b>0h</b>	<b>1h30</b>	<b>3h</b>	<b>1</b>	<b>1</b>		
<b>UET2.1.1:</b> Anglais Scientifique				1h30	3h	1	1		100%
<b>Total Semestre 2</b>	<b>336h</b>	<b>10h30</b>	<b>3h</b>	<b>10h30</b>	<b>24h</b>	<b>18</b>	<b>30</b>		

### 3 - Semestre 3

Unité d'Enseignement	VHS	V.H hebdomadaire				Coef	Crédits	Mode d'évaluation	
	14 sem	C	TD	TP	Travail personnel			Contrôle continu	Examen
<b>UE fondamentales</b>									
<b>UEF3.1</b>	<b>147h</b>	<b>4h30</b>	<b>0h</b>	<b>6h</b>	<b>9h</b>	<b>7</b>	<b>13</b>		
<b>UEF3.1.1:</b> Analyse des réseaux sociaux		1h30		1h30	3h	2	4	40 %	60 %
<b>UEF3.1.2:</b> Analyse et Exploitation de Données		1h30		1h30	3h	2	4	40 %	60 %
<b>UEF3.1.3:</b> Machine Learning Avancé		1h30		3h00	3h	3	5	40 %	60 %
<b>UEF3.2</b>	<b>147h</b>	<b>4h30</b>	<b>0h</b>	<b>6h</b>	<b>9h</b>	<b>7</b>	<b>13</b>		
<b>UEF3.2.1:</b> Big Data Analytics		1h30		3h00	3h	3	5	40 %	60 %
<b>UEF3.2.2:</b> Data Visualisation		1h30		1h30	3h	2	4	40 %	60 %
<b>UEF3.2.3:</b> Bases de données No-SQL et Cloud		1h30		1h30	3h	2	4	40 %	60 %
<b>UE méthodologies</b>									
<b>UEM3.1</b>	<b>21h</b>	<b>1h30</b>	<b>0h</b>	<b>0h</b>	<b>3h</b>	<b>1</b>	<b>2</b>		
<b>UEM3.1.1:</b> Communication et management en entreprise		1h30			3h	1	2		100 %
<b>UE transversale</b>									
<b>UET3.1</b>	<b>42h</b>	<b>1h30</b>	<b>0h</b>	<b>1h30</b>	<b>3h</b>	<b>1</b>	<b>2</b>		
<b>UET3.1.1:</b> Anglais : Méthodologie de recherche et échange		1h30		1h30	3h	1	2	40%	60 %
<b>Total Semestre 3</b>	<b>357h</b>	<b>12h</b>	<b>0h</b>	<b>13h30</b>	<b>24h</b>	<b>16</b>	<b>30</b>		

## 4 - Semestre 4

**Domaine** : Mathématiques et Informatique

**Filière** : Informatique

**Spécialité** : Data Science and Analytics

Un travail d'initiation à la recherche sera proposé à chaque étudiant. Le travail sera suivi par un enseignant et sanctionné par un mémoire et une soutenance.

Unité d'Enseignement	VHS	Coefficients	Crédits
UEF4 : Mémoire	375h	16	30
<b>Total Semestre 4</b>	375h	16	30

## 5- Récapitulatif global de la formation :

Le Volume Horaire (VH) global séparé en Cours, TD, TP, etc. pour les 04 semestres d'enseignement, pour les différents types d'UE.

<b>VH</b> \ <b>UE</b>	<b>UEF</b>	<b>UEM</b>	<b>UET</b>	<b>Total</b>
<b>Cours</b>	357h	84h	21h	462h
<b>TD</b>	105h	21h	00h	126h
<b>TP</b>	378h	42h	63h	483h
<b>Autre (PFE)</b>	375h	-	-	375h
<b>Total</b>	1215h	147h	84h	1446h
<b>Crédits</b>	105	11	4	120
<b>% en crédits pour chaque UE</b>	87.5%	9.17%	3.33 %	100%

### **III - Programme détaillé par matière**

(Une fiche détaillée par matière)

**(Tous les champs sont à renseigner obligatoirement)**

## **Intitulé du Master : Analyse et Sciences de Données**

**Semestre : S1**

**Intitulé de l'UE : UEF1.1.1**

**Intitulé de la matière : Probabilités**

**Crédits : 5**

**Coefficient : 3**

### **Objectifs de l'enseignement :**

Ce cours a pour objectif de permettre à toute étudiant disposant d'un bagage mathématique minimal de s'initier aux outils classiques de modélisation en probabilités.

### **Connaissances préalables recommandées :**

Analyse combinatoire, Calcul de probabilité, Analyse 1, Analyse 2, Algèbre.

### **Contenu de la matière :**

- Rappels sur la notion de variable aléatoire, les probabilités conditionnelles (Bayes), l'indépendance, les lois classiques et les schémas de modélisation correspondants. Inégalités classiques (Markov, Tchebychev, Hölder), lemme de Borel-Cantelli.
- Notions de convergence de variables aléatoires. Loi des grands nombres, théorème de la limite centrale.
- Vecteurs aléatoires. Vecteurs gaussiens. Théorème de Cochran.
- Conditionnement : espérance conditionnelle, lois conditionnelles, cas gaussien.
- Chaînes de Markov à temps discret et espace d'états fini ou dénombrable : définition, propriété de Markov, irréductibilité, récurrence, récurrence positive, périodicité, loi stationnaire, classification des états. Comportement asymptotique. Exemples d'applications.
- Éléments de modélisation aléatoire (processus de Poisson, ou mouvement brownien, ou processus de renouvellement...).

**Mode d'évaluation :** Examen (60%), contrôle continu (40%).

### **Références :**

- D. Foata, A. Fuchs, Processus Stochastiques, Dunod, 2004
- Karlyn, S and H. Taylor, A First Course in Stochastic Process, San Diego, 1975
- Grimmett, C; Stirzaker, D, Probability and Random Process, Oxford University Press, third edition, Oxford, 2001
- Ross, S. Introduction to Probability Models, Academic Press, seventh edition, San Diego, 2000.



**Intitulé du Master : Analyse et Sciences de Données**

**Semestre : S1**

**Intitulé de l'UE : UEF1.1.2**

**Intitulé de la matière : Mathématiques pour l'ingénierie**

**Crédits : 4**

**Coefficients : 2**

**Objectifs de l'enseignement :**

Présenter des outils mathématiques pour la modélisation et la résolution de problèmes d'optimisation.

**Connaissances préalables recommandées :**

Base d'algèbre linéaire et d'analyse (calcul matriciel, calcul différentiel et calcul d'intégrales simples)

**Contenu de la matière :**

- Algèbre bilinéaire : matrices symétriques, formes bilinéaires et produits scalaires.
- Calcul dans  $\mathbb{R}^n$  : généralités sur les fonctions de plusieurs variables, courbes de niveaux, gradient, hessien, points singuliers.
- Formules de Taylor à l'ordre  $n$ , développements limités, applications
- Analyse de Fourier : introduction aux séries de Fourier, applications.

**Mode d'évaluation :** Examen (60%), contrôle continu (40%).

**Références :**

- Mathématiques pour les sciences de l'ingénieur, F. Bertrand, M. Maumy-Bertrand, S. Ferrigno, A. Muller-Gueudin, D. Marx, Edition Dunod

## **Intitulé du Master : Analyse et Sciences de Données**

**Semestre : S1**

**Intitulé de l'UE : UEF1.1.3**

**Intitulé de la matière : Statistique exploratoire**

**Crédits : 4**

**Coefficients : 3**

### **Objectifs de l'enseignement**

Le but de ce cours est d'assurer une bonne maîtrise de l'analyse multi variée via les méthodes factorielles classiques dans des contextes de statistique exploratoire.

### **Connaissances préalables recommandées :**

Manipulation des sommes, algèbre linéaire : produit matriciel, diagonalisation

### **Contenu de la matière :**

- Méthodes de statistique multidimensionnelle pour décrire et réduire la dimension de l'espace de représentation.
- Méthodes factorielles (analyse en composantes principales, analyse factorielle discriminante, analyse factorielle des correspondances simples et multiples, analyse des corrélations canoniques)
- Analyse factorielle fonctionnelle.

**Mode d'évaluation :** Examen (60%), contrôle continu (40%).

### **Référence :**

- Bouroche JM. & Saporta G. L'analyse des données. Puf. 2010
- Tenenhaus, M. Méthodes statistiques en gestion. Dunod Entreprise. 1996
- Saporta G. , Probabilités et analyse de données en statistiques. Editions Technip. 2011
- Volle, M. Analyse des données. Economica. 1997

**Intitulé du Master : Analyse et Sciences de Données**

**Semestre : S1**

**Intitulé de l'UE : UEF1.2.1**

**Intitulé de la matière : Apprentissage et Complexité**

**Crédits : 5**

**Coefficients : 3**

**Objectifs de l'enseignement :**

Comprendre la notion d'apprentissage statistique supervisé et la notion de modèle complexe.

**Connaissances préalables recommandées : Probabilités.**

**Contenu de la matière :**

- Classification supervisée : définition de l'erreur d'ajustement et erreur de généralisation
- Notion de sur-ajustement et sparsité
- Théorie de Vapnik-Chervonenkis et entropie
- Pénalités liées à la complexité du modèle
- Extension au cadre de la régression et sélection de modèle

**Mode d'évaluation : Examen (60%), contrôle continu (40%).**

**Références**

- Statistical Learning (Hastie & Tibshirani)

## **Intitulé du Master : Analyse et Sciences de Données**

**Semestre : S1**

**Intitulé de l'UE : UEF1.2.2**

**Intitulé de la matière : Machine Learning**

**Crédits : 5**

**Coefficients : 3**

### **Objectifs de l'enseignement :**

L'objectif de ce cours est d'initier les étudiants à la théorie et à la pratique de l'apprentissage statistique (machine learning). L'accent est mis surtout sur les enjeux et les méthodes élémentaires ; les algorithmes nécessitant des outils mathématiques avancés seront vus ultérieurement.

Les TP auront pour but de tester différents algorithmes sur des problèmes simples d'apprentissage à partir de logiciels dédiés (Weka, Orange, Scikit Learn, ...) et d'implémenter certains des algorithmes vus en cours (en Python et en R).

### **Connaissances préalables recommandées :**

Notions élémentaires de probabilités et de statistiques, programmation python.

### **Contenu de la matière :**

- Présentation des enjeux et des principaux problèmes
- Apprentissage de règles
- Algorithmes élémentaires de classification
  - Supervisée : CART, plus proches voisins, SVM, Réseaux de neurones (principes, rétro-propagation)
  - Non supervisée : K-means, classification hiérarchique ascendante, méthodes basées densité, méthodes basées grille, Expectation-Maximisation.
- Evaluation d'un algorithme, courbe ROC
- Méthodes de calibration : échantillon de test, validation croisée
- Outils et workflow

**Mode d'évaluation :** Examen (60%), contrôle continu (40%)

### **Références**

- Hastie, Tibshirani, and Friedman's: The Elements of Statistical Learning
- Antoine Cornuéjols and Laurent Miclet : Apprentissage artificiel - Concepts et algorithmes

**Intitulé du Master : Analyse et Sciences de Données**

**Semestre : S1**

**Intitulé de l'UE : UEF1.2.3**

**Intitulé de la matière : Data warehouse et OLAP**

**Crédits : 4**

**Coefficients : 2**

**Objectifs de l'enseignement :**

Concevoir un système décisionnel à base d'entrepôts et magasins de données.

**Connaissances préalables recommandées :**

Modèle conceptuel de données, Bases de données relationnelles, SQL, algorithmique et programmation.

**Contenu de la matière :**

- Marché du décisionnel/Outils.
- Problématique des Entrepôts de données.
- Entrepôts de données et Magasins de données.
- Modélisation multidimensionnelle.
- Modélisation logique R-OLAP.
- Modélisation physique.
- Conception d'une solution décisionnelle.

**Mode d'évaluation :** Examen (60%), contrôle continu (40%).

**Références**

- The Data Warehouse Toolkit, R. Kimball and M. Ross.
- OLAP Solutions : Building Multidimensional Information Systems, E. Thomsen.

## **Intitulé du Master : Analyse et Sciences de Données**

**Semestre : S1**

**Intitulé de l'UE : UEM1.1.1**

**Intitulé de la matière : Marketing, techniques de communication, aspects juridiques**

**Crédits : 2**

**Coefficients : 1**

### **Objectifs de l'enseignement :**

L'objectif de cette UE est de sensibiliser, au-delà des compétences techniques sur la gestion des données, aux questions afférentes : sécurité des données, problématiques juridiques, valorisation marketing.

En outre, les cours de communication poursuivent le développement chez les étudiants d'un savoir-être interpersonnel, en se focalisant progressivement sur la gestion de projet et le management d'équipe.

**Connaissances préalables recommandées : Communication.**

### **Contenu de la matière :**

- **Communication** : développement personnel : découverte de soi, gestion du temps, styles de management, management et animation d'une équipe, conduite de réunion. Les étudiants devront aussi être capables de valoriser leurs talents et potentiels de travail lors d'un bilan de compétences, d'une réponse à des offres de stage ou d'emploi...
- **Sécurité** : l'étude des problèmes liés à l'intégrité des bases de données (BD) actives puis d'écrire les mécanismes que fournit un système de gestion de BD (SGBD) pour assurer la cohérence des données ; étudier les protocoles d'accès à la BD permettant d'assurer la confidentialité des données selon un schéma défini dans les applications métiers, apprendre à écrire des programmes d'application qui garantissent la cohérence de la BD dans un environnement concurrentiel, étudier les mécanismes qui permettent la reprise de fonctionnement normal d'une BD après la survenue d'une panne. Les aspects juridiques relatifs aux données seront également évoqués à travers des conférences d'intervenants spécialistes de la thématique.

**Mode d'évaluation : Examen (60%), contrôle continu (40%).**

### **Références**

- IT Governance : An International Guide to Data Security and ISO27001 / ISO27002 - Alan Calder et Steve Watkins
- Traitement des données personnelles, guide juridique - de Fabrice Mattatia

**Intitulé du Master : Analyse et Sciences de Données**

**Semestre : S1**

**Intitulé de l'UE : UET1.1.1**

**Intitulé de la matière : Anglais pour la communication professionnelle**

**Crédits : 1**

**Coefficients : 1**

**Objectifs de l'enseignement :**

Cette matière a pour but d'entraîner les étudiants à communiquer en employant la langue anglaise à l'université en guise de les préparer pour le monde de la recherche et de l'entreprise.

**Connaissances préalables recommandées :**

Niveau de base de la langue anglaise acquis en Licence.

**Contenu de la matière :**

- On aborde sous forme de conférence et quiz le droit en informatique (CNIL, RGPD, etc.).
- Apprentissage de la langue anglaise.
- Apprendre à s'exprimer en public, entretien, préparer son CV, etc.

**Mode d'évaluation : Examen (100%).**

**Références**

- Le droit de l'informatique, 2017, Yves Bismuth, édition l'Harmattan.
- L'Anglais au bureau, Accueillir, correspondre, téléphoner, animer des réunions, présenter, vendre..., Corinne Touati, éditeur Lgf.

## **Intitulé du Master : Analyse et Sciences de Données**

**Semestre : S2**

**Intitulé de l'UE : UEF2.1.1**

**Intitulé de la matière : Statistique Inférentielle**

**Crédits : 4**

**Coefficients : 2**

### **Objectifs de l'enseignement :**

Le but de ce cours est de consolider la maîtrise à la fois théorique et opérationnelle des notions de statistiques inférentielles de base (on se limitera en particulier au cadre iid), indispensables pour l'étude des modèles statistiques avancées, et en particulier les notions d'intervalles de confiance et de tests.

### **Connaissances préalables recommandées :**

Variations aléatoires discrètes et continues, convergence en probabilité et en loi, loi des grands nombres, théorème central limite

### **Contenu de la matière :**

- Le modèle statistique
- Estimation
- Intervalles de confiance
- La problématique d'un test paramétrique
- Les tests paramétriques classiques (z-test, t-test, f-test)
- Quelques tests non paramétriques (chi<sup>2</sup> d'adéquation ou d'indépendance, tests de rangs)

**Mode d'évaluation :** Examen (60%), contrôle continu (40%).

### **Références**

- E.L. Lehmann, Testing Statistical hypothesis, 1997, Springer.
- D. Dupont, Théorie de la décision statistique, 1986, SMG éditions.
- W. Wertz, Statistical density estimation, A survey, 1978, Vandenhoeck & Ruprecht in Gottingen.
- J. P. Florens, V. Marimoutou, A. Péguin-Feissolle, Econométrie : Modélisation et inférence, 2004, Armand Colin.
- Fomby , Hill , *Applying Kernel and Nonparametric Estimation to Economic Topics*, 2000, Advances in Econometrics.



## **Intitulé du Master : Analyse et Sciences de Données**

**Semestre : S2**

**Intitulé de l'UE : UEF2.1.2**

**Intitulé de la matière : Calcul intensif**

**Crédits : 4**

**Coefficients : 2**

### **Objectifs de l'enseignement :**

Cette matière enseigne la pratique d'agrégation de la puissance de calcul d'une manière qui offre des performances beaucoup plus élevées que celles que l'on pourrait obtenir d'un ordinateur de bureau ou d'une station de travail typique, et ce, afin de résoudre des problèmes contenant des données massifs.

### **Connaissances préalables recommandées :**

Algorithmique et programmation de niveau de Licence , Connaissances de C++

### **Contenu de la matière :**

1. Concepts de traitement parallèle (aperçu rapide)
  - a) Niveaux de parallélisme (instruction, transaction, tâche, thread, mémoire, fonction)
  - b) Modèles (SIMD, MIMD, SIMT, SPMD, modèles de flux de données, calcul basé sur la demande, etc.)
  - c) Architectures: architectures superscalaires à l'échelle N, multicœurs, multithread
2. Programmation parallèle avec CUDA
  - a) Architecture de processeur, interconnexion, communication, organisation de la mémoire et modèles de programmation dans les architectures de calcul haute performance.
  - b) Hiérarchie de la mémoire et conception de la mémoire spécifique à la transaction
3. Problèmes fondamentaux de conception dans le calcul parallèle
  - a) Synchronisation et Planification et partitionnement des tâches
  - b) Analyse des dépendances
  - c) Mappage d'algorithmes parallèles sur des architectures parallèles
  - d) Analyse des performances des algorithmes parallèles
4. Limitations fondamentales du calcul parallèle
  - a) Limitations de bande passante et limitations de latence
  - b) Techniques de masquage / tolérance de latence et leurs limites
5. Informatique et communication axée sur la puissance
  - a) Techniques de traitement sensibles à la puissance et
  - b) Conception de mémoire sensible à l'alimentation
  - c) Gestion de l'alimentation logicielle
6. Sujets avancés
  - a) Petascale Computing
  - b) Optique dans le calcul parallèle
  - c) Ordinateurs quantiques
  - d) Développements récents en nanotechnologie et son impact sur le HPC

**Mode d'évaluation :** Examen (60%), contrôle continu (40%).

### **Références**

- Bell, C.G. and Newell, A. Computer Structures. McGraw-Hill, New York 1971
- Bailey, D.H. and Buzbee, W. Private communication
- Bell, C.G. Multis: A new class of multiprocessor computers. Science 228. (Apr. 25, 1985), 452-457

**Intitulé du Master : Data Sciences and Analytics**

**Semestre : S2**

**Intitulé de l'UE : UEF2.1.3**

**Intitulé de la matière : Optimisation**

**Crédits : 5**

**Coefficients : 3**

**Objectifs de l'enseignement :**

Être capable de modéliser certains problèmes réels d'ordre économique et de se familiariser avec des méthodes de résolution

**Connaissances préalables recommandées :**

Algèbre linéaire, calcul différentiel.

**Contenu de la matière :**

- Optimisation non linéaire sans contraintes
- Conditions d'optimalités
- Méthodes de descente
- Optimisation non linéaire avec contraintes

**Mode d'évaluation :** Examens écrits (60%), contrôle continu (40%).

**Références**

- Bierlaire M., Introduction à l'optimisation différentiable, Presse Polytechniques et Universitaires Romandes, 2006.

## **Intitulé du Master : Analyse et Sciences de Données**

**Semestre : S2**

**Intitulé de l'UE : UEF2.2.1**

**Intitulé de la matière : Data Mining**

**Crédits : 5**

**Coefficients : 3**

### **Objectifs de l'enseignement :**

L'objectif de ce cours est double. Dans un premier temps (après avoir introduire le processus d'extraction de connaissance), il s'agira de présenter les méthodes de combinaison de classifieurs de la classification supervisée (les plus populaires) pour accroître la qualité de la prédiction. La seconde partie du cours sera consacrée à l'adaptation des techniques de machine learning étudiées jusqu'ici à un contexte de plus en plus populaire : le Web. Nous verrons que l'analyse de données Web regroupe 3 aspects bien différents (l'analyse du Web en tant que réseau, l'analyse de son contenu et l'analyse de ses usages) et que chacun oblige à repenser en partie les approches classiques d'apprentissage.

### **Connaissances préalables recommandées :**

Algorithmique, Statistique, Machine Learning, Bases de données, Bases de la théorie des graphes.

### **Contenu de la matière :**

- Introduction au DM (Processus KDD)
- Prétraitement de données (Nettoyage, transformation, réduction, ...), tâches et techniques du DM, évaluation de modèles et visualisation des résultats
- Recherche des modèles fréquents
- Algorithmes de classification et de prédiction plus poussées
  - **Classification supervisée:** Agrégation de classeurs et combinaison d'experts (super-learning), bagging, boosting, Random Forest.
  - **Prédiction :** Régression linéaire, régression non-linéaire, autres méthodes
- Graphe Mining : Structure de graphes, notions de centres de graphes, de clusters de graphes et de plus courts chemins.
- Web Mining : Fouille de structure (caractérisation d'un nœud dans un graphe (prestige, centralité, popularité), détection de communauté, extractions de sous-graphes sous contraintes), Fouille de contenu (classification de données textuelles, fouille d'opinion, analyse de sentiments, détection de contenu spam), Fouille d'usage (identification de sessions utilisateurs, clustering d'utilisateurs/de documents, détection de fraudes)

**Mode d'évaluation :** Examens écrits (60%), contrôle continu (40%).

### **Références**

- Han, Jiawei, Jian Pei, and Micheline Kamber. *Data Mining: concepts and techniques*. Elsevier, 2011.
- Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*. Vol. 1. No. 10. New York: Springer series in statistics, 2001.
- Liu, Bing. *Web data mining: exploring hyperlinks, contents, and usage data*. Springer Science & Business Media, 2007.

## **Intitulé du Master : Analyse et Sciences de Données**

**Semestre : S2**

**Intitulé de l'UE : UEF2.2.2**

**Intitulé de la matière : Extraction d'Information dans les Documents textes, audio, vidéo**

**Crédits : 4**

**Coefficients : 3**

### **Objectifs de l'enseignement :**

Extraire à partir de tout support d'information (texte, image, son, vidéo), les caractéristiques de contenu et de forme

**Connaissances préalables recommandées : Méthodes de classification.**

### **Contenu de la matière :**

- Objectifs, enjeux et applications.
- Extraction d'informations à partir de textes : Préliminaires (Typologie des tâches d'extraction (entités nommées, relations, indices temporels, lexique...) - Domaines d'applications -Systèmes d'extraction de l'information : architecture de base et technologies.
- Principes et méthodes : Méthodes symboliques (expressions régulières, règles et patrons) -Méthodes supervisées (CRF, arbres de décisions),
- Extraction d'information de type audio : Présentation du signal sonore, Numérisation, Extraction de paramètres temporels et fréquentiels,
- Extraction d'information de type image et vidéo : Présentation de l'image numérique, Quantification, Extraction de paramètres liés à la couleur,
- Evaluation : Méthodologie et métriques, Campagnes d'évaluation.

**Mode d'évaluation : Examen (60%), contrôle continu (40%).**

### **Références**

- Information extraction in a retrieval context : algorithms & prospects, Springer, M. F. Moens.
- Reconnaissance automatique de la parole : Du signal à son interprétation, Dunod, Haton.
- Image numérique couleur. Dunod, Trémeau.

## **Intitulé du Master : Analyse et Sciences de Données**

**Semestre : S2**

**Intitulé de l'UE : UEM2.1.1**

**Intitulé de la matière : Principes et Méthodes des Systèmes de Bases de Données Réparties**

**Crédits : 3**

**Coefficients : 2**

### **Objectifs de l'enseignement :**

Présenter les principaux problèmes posés et d'introduire les méthodes proposées dans la conception et le développement des SGBD répartis.

### **Connaissances préalables recommandées :**

Systèmes de bases de données relationnels.

### **Contenu de la matière :**

- Introduction aux BD réparties BDR.
- Principales fonctions des SGBD répartis.
- Architecture logiciel des SGBDR.
- Approches et stratégies de conception d'une BDR.
- Principe d'évaluation de requêtes réparties.

**Mode d'évaluation :** Examen (60%), contrôle continu (40%).

### **Références**

- Principles of Distributed Database Systems, February 2011. M Tamer Ozsü et Patrick Valduriez, Springer-Verlag, 3rd, ed. 2011, ISBN-13 : 978-144198833.

## **Intitulé du Master : Analyse et Sciences de Données**

**Semestre : S2**

**Intitulé de l'UE : UEM2.1.2**

**Intitulé de la matière : Simulation pour les systèmes complexes**

**Crédits : 4**

**Coefficients : 2**

### **Objectifs de l'enseignement :**

L'enseignement dispensé dans cette matière a pour objectif de traiter les problèmes de modélisation des systèmes complexes. Ils devront acquérir une compétence leur permettant de traduire un problème posé en termes de modèle, connaître les étapes essentielles pour réaliser une simulation. Ils devront être capables de piloter un projet de mise en place d'un outil de simulation dans leur propre domaine d'activité.

À la fin de ce cours, les étudiants sauront : quelles nouvelles possibilités sont offertes par les agents intelligents et les SMA; construire des systèmes multi-agents ou sélectionner pour résoudre un problème du monde réel basé sur des concepts tels que la distribution des tâches, la communication, la coopération et la coordination des actions; utiliser la technologie des agents dans des domaines tels que la collecte d'informations sur Internet, l'aide à la décision, la gestion des flux de travail.

### **Connaissances préalables recommandées :**

Génie logiciel, connaissances en intelligence artificielle et en programmation.

### **Contenu de la matière :**

- Introduction aux systèmes complexes
- Les techniques de la modélisation d'une simulation.
- Les systèmes multi agent
- Génie logiciel orienté agent, Agents, définitions et classification des agents, systèmes multi-agents
- Protocoles de communication et d'interaction des agents
- Mécanismes et stratégies de coordination, et de négociation
- Apprentissage dans les SMA
- Plateformes d'agent et aide à la décision
- Génie logiciel, Applications industrielles des SMA, SMA dans le commerce électronique, agents d'information adaptatifs et récupération d'informations

**Mode d'évaluation :** Examen (60%), contrôle continu (40%).

### **Références**

- Multi-Agent Systems: Simulation and Applications. Adelinde M. Uhrmacher, Danny Weyns
- Modeling and Simulation of Complex Systems: A Framework for Efficient Agent-Based Modeling and Simulation

## **Intitulé du Master : Analyse et Sciences de Données**

**Semestre : S2**

**Intitulé de l'UE : UET2.1.1**

**Intitulé de la matière : Anglais Scientifique**

**Crédits : 1**

**Coefficients : 1**

### **Objectifs de l'enseignement :**

Cette matière cherche à initier l'étudiant en Master 1 à lire et comprendre des articles scientifiques en anglais, et d'acquérir les outils linguistiques afin de rédiger un article ou rapport en anglais technique.

### **Connaissances préalables recommandées :**

Bonne base en langue anglaise (parlé et écrit).

### **Contenu de la matière :**

- Techniques de communication écrite.
- Présentation de méthodes de rédaction de documents différents.
  - Article de recherche.
  - Bibliographie.
  - Ouvrage ou chapitre dans un ouvrage.
  - Rapport interne de recherche.
  - PV de réunion.
  - Une demande de recrutement.
- Technique de communication orale.

Cette partie devra se faire sous forme d'exercices pratiques où l'étudiant doit communiquer oralement dans les situations (simulées) suivantes :

  - Présenter un exposé sur un travail donné.
  - Se présenter à un groupe de personnes en vue d'un recrutement.
  - Simuler une réunion de travail, etc.

**Mode d'évaluation : Examen (100%)**

**Langue d'enseignement : Anglais**

### **Références**

- TROUILLON, Jean-Louis. Chapitre 7. Enseigner l'anglais de spécialité In : Approches de l'anglais de spécialité. Perpignan : Presses universitaires de Perpignan, 2010.
- Sun, Yu-Chih, and Yu-jung Chang. "Blogging to learn: Becoming EFL academic writers through collaborative dialogues." 2012.

## **Intitulé du Master : Analyse et Sciences de Données**

**Semestre : S3**

**Intitulé de l'UE : UEF3.1.1**

**Intitulé de la matière : Analyse des réseaux sociaux**

**Crédits : 4**

**Coefficients : 2**

### **Objectifs de l'enseignement :**

Le cours présente des méthodes mathématiques et des outils de calcul pour l'analyse des réseaux sociaux (SNA). L'objectif d'apprentissage du cours est de fournir aux étudiants les connaissances essentielles pour l'analyse des réseaux qui seront applicables aux données du monde réel

### **Connaissances préalables recommandées :**

Programmation de base, Bases en théorie de graphe, Une connaissance de base en mathématiques

### **Contenu de la matière :**

- Introduction à l'analyse des réseaux sociaux. Bases de la théorie des graphes
- Analyse descriptive du réseau, Distribution des degrés, Coefficient de clusterin, Frequent patterns, Network motifs. Cliques and k-cores
- Structure du réseau, Noeuds et arêtes, Diamètre du réseau et longueur moyenne du trajet.
- Centralités des nœuds et classement sur le réseau, Mesures de centralité des nœuds : degré, proximité et Centralité intermédiaire, Eigenvector centrality and PageRank. Algorithm HITS
- Network communities, Partitionnement de graphe , Cut metrics, Edge betweenness, Modularity clustering
- Affiliation network and bipartite graphs. 1-mode projections. Recommendation systems
- Social Diffusion. Basic cascade model. Influence maximization. Les nœuds les plus influents du réseau
- Network visualization and graph layouts. Graph sampling. Low-dimensional projections
- Social media mining , Natural language processing and sentiment mining.
- SNA dans le monde réel : FB/VK and Twitter analysis , Propriétés des grands réseaux sociaux: friends, connections, likes, re-tweets

**Mode d'évaluation :** Examen (60%), contrôle continu (40%).

### **Références**

- Eric Kolaczyk, Gabor Csardi. "Statistical Analysis of Network Data with R (UseR!)". Springer, 2014. 3. Stanley Wasserman and Katherine Faust.
- David Easley and John Kleinberg. "Networks, Crowds, and Markets: Reasoning About a Highly Connected World." Cambridge University Press 2010
- Maarten van Steen. "Graph Theory and Complex Networks. An Introduction", 2010.
- Reza Zafarani, Mohammed Ali Abbasi, Huan Liu. "Social Media Mining: An Introduction". Cambridge University Press 2014.



## **Intitulé du Master : Data Science and Analytics**

**Semestre : S3**

**Intitulé de l'UE : UEF3.1.2**

**Intitulé de la matière : Analyse et Exploitation de Données**

**Crédits : 4**

**Coefficient : 2**

### **Objectifs de l'enseignement :**

Être capable de traiter des informations brutes de type texte, audio, vidéo en vue d'en dégager des informations exploitables par les décideurs.

### **Connaissances préalables recommandées :**

Extraction d'information (M1), Méthode de catégorisation/classification (K-means, KNN, SVM, EM).

### **Contenu de la matière :**

- Objectifs, enjeux et applications
- Exploitation de textes et applications
  - Catégorisation de textes : Méthodes génératives (Naive Bayes, Bernoulli), Méthodes vectorielles (Rocchio, KNN), Méthodes basées sur l'apprentissage (SVM)
  - Classifications de textes : Méthode k-means, Méthode EM
  - Résumé automatique
  - Cadre d'application : analyse et résumé d'opinions
- Exploitation de contenus sonores et applications
  - Classification de sons (phonèmes, locuteurs) : Méthodes GMM, Méthodes HMM
  - Recherche de mots clefs
  - Transcription automatique de la parole
  - Indexation audio
- Exploitation de contenus visuels et applications
  - Analyse de la texture
  - Analyse de formes et de contours
  - Classification et recherche d'images par le contenu : vers l'indexation vidéo

**Mode d'évaluation :** Examen (60%), contrôle continu (40%).

### **Références :**

- Modèles statistiques pour l'accès à l'information textuelle, Eric Gaussier, François Yvon
- Multimodal Processing and Interaction. Audio, Video, Text, Springer, Petros Maragos
- Structured Analysis Design Technic, Michel Galinier

## **Intitulé du Master : Analyse et Sciences de Données**

**Semestre : S3**

**Intitulé de l'UE : UEF3.1.3**

**Intitulé de la matière : Machine Learning Avancé**

**Crédits : 5**

**Coefficients : 3**

### **Objectifs de l'enseignement :**

Cette matière aborde des notions avancées de l'apprentissage pour traiter des problèmes complexes en sciences de données à partir de flux de données, de l'Apprentissage incrémental et constructif, apprentissage renforcé, les réseaux de neurones complexes, l'apprentissage profond ainsi que l'Apprentissage multi-tâches et apprentissage par transfert entre domaines.

### **Connaissances préalables recommandées :**

Statistiques et probabilités, mathématiques pour l'ingénierie (calcul numérique, algèbre linéaire), introduction au machine learning (S1), programmation en python.

### **Contenu de la matière :**

- Apprentissage automatique bayésien et probabiliste (Allocation de Dirichlet latente, processus gaussiens, calcul bayésien approximatif)
- Apprentissage multiple (Manifold learning)
- Apprentissage actif
- Deep learning (Motivations, principes et concepts fondamentaux, applications, plateformes, techniques de régularisation, architectures (DNN, RBM, RNN, LSTM, Auto-encodeurs, CNN, GAN))
- Transfert learning (Feature extraction, Fine-tuning)
- Deep Generative Learning
- Fondamentaux de l'apprentissage par renforcement (RL)
- Graph embeddings
- Randomized algorithms

**Mode d'évaluation :** Examen (60%), contrôle continu (40%).

### **Références**

- Heudin, Jean-Claude. *Comprendre le deep learning: une introduction aux réseaux de neurones*. Science-eBook, 2016.
- Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.
- Nielsen, Michael A. *Neural networks and deep learning*. Vol. 2018. San Francisco, CA, USA:: Determination press, 2015.

## **Intitulé du Master : Analyse et Sciences de Données**

**Semestre : S3**

**Intitulé de l'UE : UEF3.2.1**

**Intitulé de la matière : Big Data Analytics**

**Crédits : 5**

**Coefficients : 3**

### **Objectifs de l'enseignement :**

Connaissances sur les différents types d'analyses de données massives (Big Data Analytics) que les organisations peuvent utiliser. Comment l'analyse de données massives peut aider, par exemple, à faire de la prédiction, à comprendre des pratiques, à opérer des analyses à grande échelle, etc.

### **Connaissances préalables recommandées :**

Statistiques, machine learning, data mining, outils analytiques, bases de données, curiosité intellectuelle.

### **Contenu de la matière :**

- Définir le concept de Big Data (Volume, Variété, Vitesse, etc.)
- Comprendre et illustrer la problématique de passage à l'échelle
- Comprendre le concept de valorisation de la donnée
- Comprendre les approches d'intégration de données massives
- Comprendre la problématique de traitement de données massives à des fins d'analyse distribuer (partitionner) des données sur plusieurs nœuds de calcul
- Comprendre et appliquer le paradigme Map/Reduce
- Configurer des systèmes Map/Reduce et déployer des applications
- Découvrir la palette d'outils d'analyse de données massives
- Identifier les risques et opportunités du Big Data pour les individus (vie privée, protection des données personnelles...)
- Conduire un projet Big Data en entreprise (étapes projets, bonnes pratiques, schémas organisationnels et gouvernance des données)
- Situer les notions connexes sur la visualisation de données, les objets connectés, l'Open Data etc. vis-à-vis de l'analyse de données massives

**Mode d'évaluation :** Examen (60%), contrôle continu (40%).

### **Références :**

- Kleppmann, Martin. *Designing data-intensive applications: The big ideas behind reliable, scalable, and maintainable systems.* " O'Reilly Media, Inc.", 2017.
- Davenport, Thomas H., Paul Barth, and Randy Bean. *How 'big data' is different.* MIT Sloan Management Review, 2012.
- Marz, Nathan, and James Warren. *Big Data: Principles and best practices of scalable real-time data systems.* New York; Manning Publications Co., 2015.
- Reynolds, Vince. *Big Data for Beginners: Understanding SMART Big Data, Data Mining & Data Analytics for Improved Business Performance, Life Decisions & More!.* CreateSpace Independent Publishing Platform, 2016.

## **Intitulé du Master : Analyse et Sciences de Données**

**Semestre : S3**

**Intitulé de l'UE : UEF3.2.2**

**Intitulé de la matière : Data Visualisation**

**Crédits : 4**

**Coefficients : 2**

### **Objectifs de l'enseignement :**

L'objectif de ce cours est de comprendre les principes de base visualisation d'information, et de maîtriser les outils permettant de visualiser de manière efficace des données

**Connaissances préalables recommandées :** algorithmes et programmation web de niveau de Licence et M1

### **Contenu de la matière :**

- **Principes de design:** Il s'agit de comprendre les principes permettant de représenter efficacement des données (comment encoder différentes dimensions ? avec quels effets perceptifs ? principes de Tufte, etc.). Nous critiquerons des visualisations populaires, et proposerons des alternatives. Nous nous appuierons sur le logiciel de visualisation Tableau, Power BI, QlikView.
- **Aspects mathématiques et algorithmiques:** Nous verrons que les données en hautes dimensions, ou trop nombreuses, peuvent poser des problèmes de visualisation. Nous étudierons des méthodes de clustering et réduction de dimension.
- **Visualisation interactive:** Nous verrons enfin les principes de base d'interaction. Et verrons comment créer des visualisations interactives pour le Web grâce à la bibliothèque javascript d3.js.

Les étudiants rendront par ailleurs un projet en fin de semestre. Il s'agira d'une visualisation interactive web, avec screencast et rendu d'un rapport sous forme d'article scientifique.

**Mode d'évaluation :** Examen (60%), contrôle continu (40%).

### **Références**

- Interactive Data Visualization for the Web by Scott Murray 2nd Edition (2017)
- D3.js in Action by Elijah Meeks 2nd Edition (2017)
- Semiology of Graphics by Jacques Bertin (2010)
- The Grammar of Graphics by Leland Wilkinson

## **Intitulé du Master : Analyse et Sciences de Données**

**Semestre : S3**

**Intitulé de l'UE : UEF3.2.3**

**Intitulé de la matière : Bases de données (NoSQL et Cloud)**

**Crédits : 4**

**Coefficients : 2**

### **Objectifs de l'enseignement :**

Comprendre les principaux problèmes posés et les méthodes proposées dans la conception des SGBD parallèles et les systèmes clouds. Connaître les principes des bases de données NoSQL et apprendre à les manipuler.

### **Connaissances préalables recommandées :**

Bases de données relationnelles, algorithmique et programmation

### **Contenu de la matière:**

- Objectifs des BD parallèles
- Méthodes de répartition de données
- Introduction aux traitements de données dans les systèmes clouds
- Evaluation et optimisation élastique de requêtes décisionnelles
- Etude comparative entre les SGBD parallèles et les systèmes clouds
- NoSQL
  - Panorama des différents types de bases NoSQL
  - Manipulation d'une base de données NoSQL

**Mode d'évaluation :** Examen (60%), contrôle continu (40%).

### **Références :**

- Les bases de données NoSQL : Comprendre et mettre en œuvre - Rudi Bruchez, Eyrolles
- MongoDB The Definitive Guide, Kristina Chodorow, O'Reilly
- Principles of Distributed Database Systems, T. Ozu P. Valduriez, 3rd edition, 2011
- Traitement parallèle dans les bases de données relationnelles : concepts, méthodes et applications, A. Hameur- lain, P. Bazex, F. Morvan, 1996.

**Intitulé du Master : Analyse et Sciences de Données**

**Semestre : S3**

**Intitulé de l'UE : UEM3.1.1**

**Intitulé de la matière : Communication et management en entreprise**

**Crédits : 2**

**Coefficients : 1**

**Objectifs de l'enseignement :**

- Fournir aux étudiants les concepts de base du cycle de décision dans les entreprises
- Savoir préparer et mener un entretien d'embauche professionnel.

**Connaissances préalables recommandées :**

Bonne connaissance de ce que sont les systèmes d'information.

**Contenu de la matière :**

- **Cycle de décision en entreprise** : En cycle de décision nous développons les deux principales méthodes de prise de décision et présentons l'approche systémique de l'entreprise. Nous développons les trois niveaux de la prise de décision (stratégique, tactique, opérationnel). L'intelligence économique est également présentée à la fois comme un mode de gouvernance et en tant que science. Les trois axes de l'intelligence économique sont détaillés, avec un accent particulier mis sur la veille stratégique.
- **Techniques de communication** : Ce cours aborde l'entretien d'embauche en face à face : l'importance de la préparation et du non verbal, les techniques professionnelles relevant de l'entretien, les erreurs à éviter, les outils pour suivre sa candidature.

**Mode d'évaluation : Examen (100%).**

**Références :**

- M Salles. Décision et système d'information.
- JR Alcaras. Les conceptions de la décision en sciences économiques : vers une approche ingénieriale ?

**Intitulé du Master : Data Science and Analytics**

**Semestre : S3**

**Intitulé de l'UE : UET3.1.1**

**Intitulé de la matière : Anglais : Méthodologie de recherche et échange**

**Crédits : 2**

**Coefficients : 1**

**Objectifs de l'enseignement :**

- Initiation à la rédaction scientifique en anglais
- Fournir des critères permettant d'assurer la scientificité d'une recherche
- Apprentissage du langage de rédaction Latex

**Connaissances préalables recommandées :**

Bonne base en langue anglaise (parlé et écrit).

**Contenu de la matière:**

- Histoire de la recherche scientifique:
  - Les différentes formes de Recherche Scientifique.
  - Organisation de la recherche
  - Les livrables de la recherche
  - Les Métiers de la recherche
  - Financement de la recherche scientifique.
  - Évaluation de la recherche
- Méthodologie de recherche: les différents types de recherche (théorique, expérimentale,...)
  - Démarche pour mener et présenter un travail de recherche:
  - Cheminement de l'idée au résultat,
  - Formalisation d'un travail de recherche,
  - Réponse à appel d'offre (PNR, CNEPRU, ...),
  - Différentes formes de divulgation des résultats de la recherche.
- Éthique dans la démarche de recherche:
  - Introduction : C'est quoi l'éthique scientifique ?
  - L'éthique dans les pratiques scientifique ou La fraude scientifique
  - L'expérimentation : sur les humaines et les animaux.
- Science et société : Application des travaux de recherche dans la vie quotidienne
- Le langage Latex

**Mode d'évaluation :** Examen (100%).

**Références :**

- Deiy, D., Mock, T.J. (1985), «Information Support Systems for Problem Solving», Décision Support Systems, Vol. 1, n°2, p. 103-109.
- Berg-ron, F.,Blouin, C.(1980), «L'évaluation de Systèmes d'Information», document de travail n°80-15,Faculté des Sciences de l'Administration, Université de Lavai, Québec, Canada.
- Boland, R.J., Hirschhelm, R.A. (1987). Critical Issues in Information Systems Research, John Wiley & Sons.Boland, R.J.. Tenkasi, R.V., Te'eni, D. (1994), «Designing Information Technologic to Support Distributed Cognition», Organization Science,Vol. 5, n°3, p. 457-475.