

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

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

OFFRE DE FORMATION L.M.D.

MASTER ACADEMIQUE

Etablissement	Faculté / Institut	Département
Université de Blida 1	Sciences	Informatique

Domaine	Filière	Spécialité
MI	Informatique	Data Science and Natural Language Processing

Responsable de l'équipe du domaine de formation :

Pr. Mustapha CHELLALI

الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي والبحث العلمي

عرض تكوين

ل. م. د

ماستر أكاديمي

المؤسسة	الكلية/ المعهد	القسم
جامعة البليدة 1	العلوم	الاعلام الآلي

الميدان	الشعبة	التخصص
رياضيات وإعالم آلي	الاعلام الآلي	علوم البيانات و المعالجة الآلية للغة

مسؤول فرقة ميدان التكوين :

البروفيسور مصطفى شلال



SOMMAIRE

I.	Fiche d'identité du Master.....	1
1.	Localisation de la formation.....	1
2.	Coordinateurs.....	1
3.	Partenaires extérieurs.....	1
4.	Contexte et objectifs de la formation.....	2
A.	Organisation générale de la formation : position du projet.....	2
B.	Conditions d'accès.....	3
C.	Objectifs de la formation.....	3
D.	Profils et compétences visées.....	3
E.	Potentialités régionales et nationales d'employabilité.....	4
F.	Passerelles vers les autres spécialités.....	4
G.	Indicateurs de suivi du projet.....	4
5.	Moyens humains disponibles.....	5
A.	Capacité d'encadrement.....	5
B.	Equipe d'encadrement de la formation.....	5
6.	Moyens matériels disponibles.....	10
A.	Laboratoires Pédagogiques et Equipements.....	10
B.	Terrains de stage et formation en entreprise.....	11
C.	Laboratoire(s) de recherche de soutien à la formation proposée.....	11
D.	Projet(s) de recherche de soutien à la formation proposée	
E.	Documentation disponible.....	12
F.	Espaces de travaux personnels et TIC.....	12
II.	Fiche d'organisation semestrielle des enseignements.....	13
1.	Semestre 1.....	14
2.	Semestre 2.....	15
3.	Semestre 3.....	16
4.	Semestre 4.....	17
5.	Récapitulatif global de la formation.....	17
III.	Programme détaillé par matière.....	19
IV.	Accords ou conventions.....	53
V.	Curriculum Vitae des Coordonateurs.....	58

I. Fiche d'identité du Master

1. Localisation de la formation

Faculté (ou Institut) : Sciences

Département : Informatique

Section :

2. Coordinateurs

- Responsable de l'équipe du domaine de formation

(Professeur ou Maître de conférences Classe A) :

Nom & prénom : Pr. Mustapha CHELLALI

Grade : Professeur

☎ : 025-27-24-52

E - mail : m_chellali@yahoo.com

Joindre un CV succinct en annexe de l'offre de formation (maximum 3 pages)

- Responsable de l'équipe de la filière de formation

(Maitre de conférences Classe A ou B ou Maitre Assistant classe A) :

Nom & prénom : Dr. Messaouda FAREH

Grade : Mmaître de Conférences grade -A-

☎ : 0776-48-39-73

E - mail : farehm@gmail.com

Joindre un CV succinct en annexe de l'offre de formation (maximum 3 pages)

- Responsable de l'équipe de spécialité

(au moins Maitre-Assistant Classe A) :

Nom & prénom : Dr. Melyara MEZZI

Grade : Maître de Conférences grade -B-

☎ : 0792-351-842

E - mail: mezzi_melyara@univ-blida.dz

Joindre un CV succinct en annexe de l'offre de formation (maximum 3 pages)

3. Partenaires extérieurs

Partenaires Nationaux :

- Université d'Alger 2. Département sciences du langage
- Université D'Oran. Faculté des lettres, des langues et des arts
- Académie Algérienne de la Langue Arabe (Alger)
- Centre de Recherche Scientifique et Technique pour le Développement de la
- Langue Arabe, CRSTDLA (Alger)
- Entreprises et autres partenaires socio-économiques

Partenaires internationaux :

- Centre TAL, Lucien Tesnière, Besançon (France)
- Centre TAL, Université de Wolverhampton, Grande Bretagne
- Centre TAL, Université de Barcelone (Espagne)
- INALCO (Dépt. TAL) Paris
- Réseau LTT = Linguistique Terminologie Traduction (AUPELF), Réseau Vives (Xarxa / CRUO), Espagne.

4. Contexte et objectifs de la formation

A. Organisation générale de la formation : position du projet

Si plusieurs Masters sont proposés ou déjà pris en charge au niveau de l'établissement (même équipe de formation ou d'autres équipes de formation), indiquez dans le schéma suivant, la position de ce projet par rapport aux autres parcours.

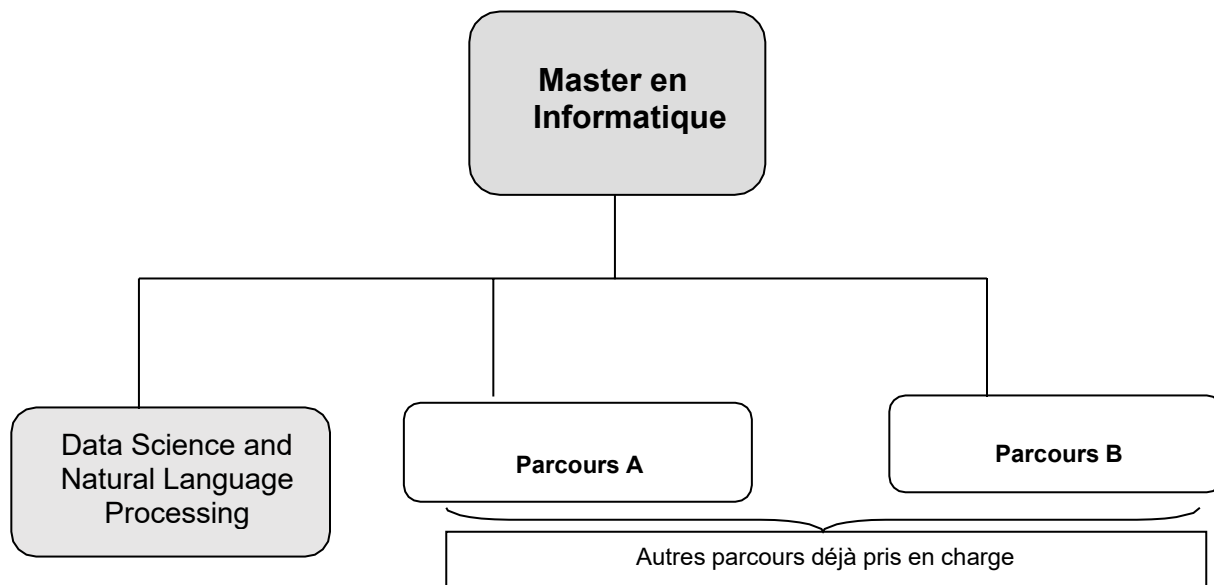
La science des données est l'étude scientifique des données pour acquérir des connaissances. Ce domaine combine plusieurs disciplines pour extraire des connaissances à partir d'ensembles de données massifs dans le but de prendre des décisions et de faire des prédictions éclairées. Les scientifiques des données, les analystes de données, les architectes de données, les ingénieurs de données, les statisticiens, les administrateurs de bases de données et les analystes commerciaux travaillent tous dans le domaine de la science des données.

Le besoin de science des données augmente rapidement, car la quantité de données augmente de façon exponentielle et les entreprises dépendent davantage de l'analytique pour générer des revenus et innover. Par exemple, les interactions commerciales devenant de plus en plus numériques, davantage de données sont créées, ce qui offre de nouvelles possibilités de tirer des enseignements sur la façon de mieux personnaliser les expériences, d'améliorer le service et la satisfaction des clients, de développer des produits nouveaux et améliorés et d'augmenter les ventes. En outre, dans le monde des affaires et au-delà, la science des données a le potentiel d'aider à résoudre certains des défis les plus difficiles du monde.

Le Traitement Automatique de la Langue Naturelle (TALN) ou Traitement Automatique des Langues (TAL), plus couramment appelé NLP (de l'anglais : Natural Language Processing) est un domaine multidisciplinaire impliquant la linguistique, les mathématiques et l'intelligence artificielle, qui vise à créer des outils de traitement de la langue naturelle pour diverses applications. En effet, le TAL est sorti des laboratoires de recherche pour être progressivement mis en œuvre dans des applications informatiques nécessitant l'intégration du langage humain à la machine.

L'objectif des data scientists est essentiellement divisé en trois domaines : *la collecte de Données*, *l'Analyse de Données* et *la Déduction d'Informations* à partir de données. Chacune de ces tâches nécessite du personnel spécialisé, prend du temps et coûte de l'argent. Pourtant, à l'ère de l'information dans laquelle nous vivons, la tâche culminante consiste à transformer les données en produits. Par conséquent, ce domaine attire l'attention de nombreux groupes de Recherche dans le milieu universitaire ainsi que dans l'industrie. Au cours des dernières décennies, les approches basées sur les données ont vu le jour et ont gagné en popularité car elles nécessitent beaucoup moins d'efforts humains.

Le traitement automatique du langage naturel (TAL) fait partie des domaines fortement influencés par les données et donc ce Master en TAL et Sciences des Données prépare des compétences dans l'Ingénierie des Langues et des données. Il s'adresse aussi bien aux étudiants issus de filières technologiques : informaticiens, mathématiciens et/ ou électroniciens désireux de se positionner comme Data Scientist impliqué dans les technologies Linguistiques dans le marché académique ou professionnel.



B. Conditions d'accès

(indiquer les parcours types de licence qui peuvent donner accès à la formation Master proposée)

Licences : Licence informatique. Licence mathématique et éventuellement Électronique.

La sélection du candidat se fait sur étude du dossier et entretien pour évaluer sa capacité à suivre la formation en master.

C. Objectifs de la formation

(Compétences visées, connaissances acquises à l'issue de la formation- maximum 20 lignes)

Pour répondre aux besoins sans cesse grandissants des entreprises et des laboratoires de recherche et de développement en spécialistes de haut niveau maîtrisant parfaitement les modèles et formalismes linguistiques pour pouvoir traiter des problématiques liées aux Sciences des données avec ses divers domaines applicatifs : Recherche d'Information, Résolution de problèmes, Data Analytics, Syntaxe et sémantique des textes, Génération et traduction automatique, etc. Cette formation a pour objectif de répondre à des besoins de formation dans le domaine de l'ingénierie des langues pour alimenter aussi bien nos différents laboratoires de recherche que ceux d'autres organismes universitaires et industriels en spécialistes dans ce domaine.

D. Profils et compétences visées :

(maximum 20 lignes) :

A l'issue de sa formation, le lauréat aura une maîtrise du domaine du Traitement Automatique de la Langue et des Sciences des Données. Le lauréat formé aura des connaissances dans toutes les domaines applicatifs liés au TAL. Son profil lui offrira un fort

degré d'employabilité dans plusieurs secteurs et lui permet d'entrer de plain-pied aussi bien dans les secteurs économiques que ceux de recherche.

E. Potentialités régionales et nationales d'employabilité

Retombées :

- *Former des cadres de haut niveau maîtrisant le domaine du Traitement automatique de la langue et le Datascience, notamment les formalismes et modèles appliqués aux langues.*

Débouchées :

- **Académiques :** Ce Master mène tout naturellement à la préparation d'un travail de doctorat en Traitement automatique de la langue & Data Science.

- **Professionnels :**
 - La formation dispensée prépare les étudiants à des travaux aussi bien de recherche que de développement dans tous les secteurs industriels qui font appel aux notions inculquées et compétences acquises dans le domaine.
 - Par ailleurs, Les centres de recherche tels que le CERIST, le CDTA et le CRSTDLA ainsi que l'AALA ont des besoins à couvrir pour alimenter les divisions de recherche travaillant dans le domaine des industries de la langue.
 - Pour finir, plusieurs compagnie privée et startups voient le jour en Algérie. Le but de ces dernières est de profiter des apports incontestés de l'alliance des deux domaines couverts par la formation afin d'apporter des solutions, des services, ou répondre à des besoins en information, en digital-image, assistance...etc.

F. Passerelles vers les autres spécialités

Cette formation, pourrait se positionner comme passerelle ou spécialité pivot à d'autres spécialités. A titre d'exemple :

- Intelligence Artificielle
- Système Intelligents d'aide à la décision
- Systèmes embarqués intelligents
- Internet of Things
- Psychologie Cognitive
- Linguistique Computationnelle
- Data Analytics
- Big Data

G. Indicateurs de suivi du projet

- Examens de moyenne durée
- Contrôle continu
- Travail personnel
- Comités Pédagogiques Modulaires
- Comités Pédagogiques de Coordination
- Comités et Conseils Scientifiques

5. Moyens humains disponibles

A. Capacité d'encadrement

(exprimé en nombre d'étudiants qu'il est possible de prendre en charge) :

Suite aux expériences précédentes avec la formation TAL, la capacité d'encadrement dans cette discipline peut atteindre soixante (60)étudiants.

B. Equipe d'encadrement de la formation

B-1 : Encadrement Interne

Nom, prénom	Diplôme	Grade	Laboratoire de recherche de rattachement	Type d'intervention *	Emargement
BENBLIDIA Nadjia	Doctorat	Professeur	LRDSI	Séminaires	
BACHA Siham	Doctorat	MCB	Design Smart systems, LATSI, Université de Blida 1	Cours, TP	
CHERIGUENE Soraya	Doctorat	MCB	LabGED, Université BADJI Mokhtar Annaba	Cours, TP	
DJEMIA Nora	Doctorat	MCB	LRDSI	Cours, TD	
GHEBGHOUB Yasmine	Doctorat	MCB	LRDSI	Cours, TD, TP	
HIRECHE Célia	Doctorat	MCB	LRIA, USTHB	Cours, TD, TP	
KAMECHE Abdellah	Doctorat	MAA	LRDSI	Cours, TD, TP	
MEZZI Melyara	Doctorat	MCB	LRDSI	Cours, TD, TP	
OUKID Lamia	Doctorat	MCB	LRDSI	Cours, TD, TP	
OUKID Saliha	Doctorat	Professeur	LRDSI	Cours, TD, TP	
RIALI Ishak	Doctorat	MCB	LRDSI	Cours, TD, TP	
ZAHRA Fatima Zohra	Magister	MAA	LRDSI	Cours, TD, TP	
Rahiani Nessrine	Doctorat	MCB	LRDSI		
BEY Fella	Doctorat	MCB	LRDSI		
DAOUD Hayat	Doctorat	MCB	LRDSI		
BENYAHIA Nadjia	Magister	M.A.A.	LRDSI		
Haraj Ikram		M.A.A.			

* = Cours, TD, TP, Encadrement de stage, Encadrement de mémoire, autre (à préciser)

Nom, prénom	Diplôme	Grade	Laboratoire de recherche de rattachement	Type d'intervention *	Emargement
BALA Mahfoud	Doctorat	MC-B		Cours / TD / TP	
DOUGA YASSINE	DOCTORAT	MCB	LRDSI		
FERFERA SOUFIANE	MAGISTER	MAA	LRDSI		
Chenguerre Selma	Doctorat	MCB	LabGED, Annaba	Cours / TP	
BERRANDANE Djannet	Magister	MAA	EDIS Sidi Bel Abbes	Cours TD / TP	
Midoun Khedidja	Doctorat	MCB	LRDSI	Cours / TD / TP	
Zatra Fatma Zohra	Doctorat	MAA	LRDSI	Cours / TD / TP	
BOUTOUINI Bachir	Magister	MAA	LRDSI	Cours / TD / TP	
OUAMPANI Wale	Magister	MAA	LRDSI	Cours / TD / TP	
DUL-D-KH AOUA Nel	Doctorat	PROFESSEUR	LRDSI	Cours / TD / TP	
FAREH Merguenda	Doctorat	MCA	LRDSI	Cours / TD / TP	
Boumrahli Fatma	DOCTORAT	MCA	LRDSI	Cours / TD / TP	

Avis du chef du département d'Informatique

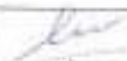
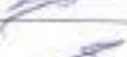


Avis du responsable de l'équipe du domaine de formation



B-2 : Encadrement Externe

Nom, prénom	Diplôme	Etablissement de rattachement	Type d'intervention *	Emargement
Pr. BOUHADIBA Farouk A.N.	PhD	Linguistique, Dynamique du Langage et Didactique (LDD)-Université d'Oran	Séminaires, Cours + TD Encadrement	
Pr. CARDEY Sylviane	Doctorat de 3ème cycle 1984, Doctorat d'Etat 1987	Centre TAL Lucien Tesnière (France)		
Pr. GREENFIELD Peter	MA (Cantab), PhD (Birmingham: Computer Science), HDR	Centre TAL Lucien Tesnière (France)		
Pr. POGNON Patrice	Docteur d'Etat es Lettres (Université de Paris 3, 1977)	INALCO (France) TAL		
Pr. MITKOV Ruslan		Wolverhampton University, UK TAL		
Pr. BLANCO Xavier		Universitat Autònoma de Barcelona (Espagne) TAL		
Dr Aliane Hassina	Doctorat d'état	CERIST	Séminaires, Cours + TD Encadrement	
Pr Hadj Salah Abderrahmane	Doctorat d'état	Université Alger 2	Séminaires,	

Pr salmi Abdelmadjid	Doctorat d'état	Université d'Alger 2	Séminaires, Cours + TD	
Fatiha KHELOUT	Doctorat en science	AALA-CRSTDLA	Séminaires, Cours + TD	
Fouzia BADAOU	Doctorat en science	CRSTDLA	Séminaires, Cours + TD	
Ghania DROUA	Doctorat en science	CRSTDLA	Séminaires, Cours + TD	
Abbas Mourad	Doctorat en science	CRSTDLA	Séminaires, Cours + TD	

* = Cours, TD, TP, Encadrement de stage, Encadrement de mémoire, autre (à préciser)

B-3 : Synthèse globale des ressources humaines

Grade	Effectif Interne	Effectif Externe	Total
Professeurs	2	8	9
Maîtres de Conférences (A)	0		0
Maîtres de Conférences (B)	7	4	11
Maître Assistant (A)	1		1
Maître Assistant (B)			
Autre (préciser)			
Total	10	12	22

B-4 : Personnel permanent de soutien

(indiquer les différentes catégories)

Grade	Effectif
Secrétaire	1
Technicien	3
Ingénieur	1

6. Moyens matériels disponibles

A. Laboratoires Pédagogiques et Equipements

Fiche des équipements pédagogiques existants pour les TP de la formation envisagée (1 fiche par laboratoire)

Intitulé du laboratoire : TP Réseaux, Sécurité Réseaux

Capacité en étudiants : 24

N°	Intitulé de l'équipement	Nombre	Observations
01	Câbles de paires torsadées souple	500m	
02	Câbles de paires torsadées rigide	2000m	
03	Fibre optique Monomode	1000m	
04	Connecteurs RJ45	500	
05	Connecteurs FO ST/SC	50	
06	Pince à sertir RJ45/RJ11	5	
07	Valise pour Collage Fibre Optique	3	
08	Valise pour Sertissage Paire Torsadée	1	
09	Fusionneuse de Fibre Optique	1	
10	PABX 4/16	1	
11	Modem Analogique externes	4	
12	Convertisseurs 100Base FX/100 Base TX	3	
13	Convertisseurs 100Base LX/100 Base TX	3	
14	Testeurs Connexion RJ45	1	
15	Prises murales RJ45	20	
16	Compatible PC + Onduleurs (en réseau)	24	
17	Serveur	1	

B. Terrains de stage et formation en entreprise

Lieu du stage	Nombre d'étudiants	Durée du stage
Département d'Informatique, U. Blida 1	20	03 Mois
CRSTDLA, Bouzareah, Alger	20	03 Mois
AALA, Alger	20	03 Mois
CDTA, Baba Hassen, Alger	10	03 Mois
CERIST, Alger	10	03 Mois

C. Laboratoire(s) de recherche de soutien à la formation proposée

Laboratoire(s) de recherche de soutien au master :

Chef du laboratoire

N° Agrément du laboratoire

Avis du chef de laboratoire :

Université Saad Dahlab de Blida
Laboratoire de Recherche Pour le
Développement des Systèmes Informatiques

Date :

26/02/2024

[Signature]

D. Projet(s) de recherche de soutien à la formation proposée

Intitulé du projet de recherche	Code du projet	Date du début du projet	Date de fin du projet
Représentation et raisonnement des connaissances à partir des données liées	C00L07UN090120220001	Janvier 2022	Décembre 2025
Architectures neuronales pour la détection visuelle/textuelle du cancer du sein à partir des données massives	C00L07UN090120220002	Janvier 2022	Décembre 2025
Sécurité des réseaux sociaux	C00L07UN090120220003	Janvier 2022	Décembre 2025

E. Documentation disponible

(en rapport avec l'offre de formation proposée)

La bibliothèque de faculté et la bibliothèque centrale disposent de livres sur le Traitement Automatique de la Langue et les Data Sciences et sur d'autres domaines connexes de ce master. Aussi, les compte SNDL et Iqraa permettent de nos jours l'accès à un large éventail de livres, manuels, articles et revues scientifiques de renommées internationales.,

De plus, dans le cadre du partenariat avec le laboratoire LRDSI, des comptes professionnels, permettant l'accès à des ressources logicielles, services cloud, cours en lignes, peuvent être envisagés dans les plateformes suivantes :

- Datacamp
- PluralSight
- LinkedIn Learning
- Udacity
- Coursera
- EdX
- DataQuest, etc.

F. Espaces de travaux personnels et TIC

Plusieurs salles disposants d'ordinateurs connectés à l'internet (réseaux du MESRS) sont disponibles au niveau du département et au niveau du Centre de Calcul et de la bibliothèque centrale de l'université (une dizaine de salles d'une capacité totale de cents postes).

II. Fiche d'organisation semestrielle des enseignements

(Prière de présenter les fiches des 4 semestres)

1. Semestre 1

Unité d'Enseignement	VHS	V.H hebdomadaire				Coeff	Crédits	Mode d'évaluation	
	14-16 sem	C	TD	TP	Autres			Continu	Examen
UE fondamentale									
UEF(O/P) :									
Programmation pour la Data Science	45h00	1h30		1h30		3	6	60%	40%
Bases de données avancées	67h30	1h30	1h30	1h30		3	6		
Algorithmique avancée et Calcul de complexité	67h30	1h30	1h30	1h30		3	6	60%	40%
UE méthodologie									
UEM(O/P) :									
Computational Linguistics	67h30	1h30	1h30	1h30		2	3	60%	40%
Théorie des langages formels	45h00	1h30	1h30			2	3	60%	40%
Traitement des données Linguistiques	45h00	1h30		1h30		2	3	60%	40%
UE découverte									
UED(O/P)									
Mathématiques appliquées aux sciences de données	22h30	1h30	1h30			1	2		100%
UE transversales									
UET(O/P)									
Sécurité des Systèmes Informatiques	22h30	1h30				1	1		100%
Total Semestre 1	360h00	12h00	7h30	6h30		17	30		

2. Semestre 2

Unité d'Enseignement	VHS	V.H hebdomadaire				Coeff	Crédits	Mode d'évaluation	
	14-16 sem	C	TD	TP	Autres			Continu	Examen
UE fondamentales									
UEF1(O/P) :									
Apprentissage automatique et applications neuronales	45h00	1h30		1h30		3	6	60%	40%
Data mining	67h30	1h30	1h30	1h30		3	6	60%	40%
Base de Données NoSQLpour le Big Data	45h00	1h30		1h30		3	6	60%	40%
UE méthodologie									
UEM(O/P)									
Principes de conception et inférence causale pour la prise de décision basée sur les données	45h00	1h30	1h30			2	3	60%	40%
Entrepôts de données et Analyse OLAP	22h30	1h30				2	3		100%
Algorithmique avancée et Résolution de problèmescomplexes	67h30	1h30	1h30	1h30		2	3	60%	40%
UE de découverte									
UED(O/P)									
Éthique du travail	22h30	1h30				1	2		100%
UE transversale									
UET(O/P)									
Business Intelligence	22h30			1h30		1	1		100%
Total Semestre 2	360h00	10h30	4h30	7h30		17	30		

3. Semestre “

Unité d'Enseignement	VHS	V.H hebdomadaire				Coeff	Crédits	Mode d'évaluation	
	14-16 sem	C	TD	TP	Autres			Continu	Examen
UE fondamentales									
UEF1(O/P)									
Big data Analytics	67h30	1h30		1h30		3	6	60%	40%
Algorithmique et modélisation linguistique	45h00	1h30		1h30		2	4	60%	40%
Méthodes chronologiques et prévision	45h00	1h30		1h30		2	4	60%	40%
UE méthodologie									
UEM(O/P)									
Recherche d'information	45h00	1h30		1h30		2	3	60%	40%
Graphes de connaissances et apprentissage	67h30	1h30	1h30	1h30		2	3		100%
Extraction, analyse et visualisation des données	45h00	1h30		1h30		2	3	60%	40%
UE découverte									
UET1(O/P)									
Entrepreneuriat	22h30	1h30				1	2		100%
UE Transversale									
UED1(O/P)									
Techniques de rédaction	22h30	1h30				1	1		100%
Total Semestre 3	360h00	12h00	1h30	09h00		17	30		

4. Semestre 4

Domaine : Mathématiques et Informatique
Filière : Informatique
Spécialité : Data Science and Natural Language Processing

Stage en entreprise sanctionné par un mémoire et une soutenance.

	VHS	Coeff	Crédits
Travail Personnel	600 H		
Stage en entreprise	600 H	1	30
Séminaires			
Autre (préciser)			
Total Semestre 4	1200 H	1	30

5. Récapitulatif global de la formation

(Indiquer le VH global séparé en cours, TD, pour les 04 semestres d'enseignement, pour les différents types d'UE)

VH \ UE	UEF	UEM	UED	UET	Total
Cours	315	135		82h30	525
TD	180	22h30			202h30
TP	285 h	90			382h30
Travail personnel	1420	45	45	22h30	1532h30
Autre (préciser)	600				600
PFE					
Total	2800h	292h30	45	105h	3241h30
Crédits	90	21		9	120
% en crédits pour chaque UE	60%	30%	6.7%	3.3%	100 %

III. Programme détaillé par matière (1 fiche détaillée par matière)

Intitulé du Master : DATA SCIENCE and NATURAL LANGUAGE PROCESSING

Semestre : S1

Intitulé de l'UE : UEF1.1

Intitulé de la matière : Programmation pour la Data Science

Crédits : 6

Coefficients : 3

Objectifs de l'enseignement (*Décrire ce que l'étudiant est censé avoir acquis comme compétences après le succès à cette matière – maximum 3 lignes*).

Ce module vise à fournir aux étudiants une compréhension approfondie des concepts du langage de programmation Python afin de pouvoir faciliter la manipulation des données et le développement des applications Python.

Au terme de ce cours, l'étudiant(e) sera capable de :

- *Comprendre les fondamentaux du langage ;*
- *Créer et manipuler des structures de données ;*
- *Utiliser des modules standards et créer ses propres modules ;*
- *Développer et maintenir des applications Python.*

Connaissances préalables recommandées

L'étudiant doit avoir des connaissances de base en programmation.

Contenu de la matière

I. Concepts fondamentaux de python

- Syntaxe et opérations de base
- Structure de données primitives (integer, float, ..., etc.)
- Structure de données non primitives (set, dictionnaire, tuple, ..., etc.)
- Les fonctions et programmation fonctionnelle avec map, reduce, filter et fonction anonyme
- Les fichiers, les erreurs et les exceptions

II. Manipulation des données avec Pandas

- Introduction aux dataframes, Series, Index Objects
- Fonctionnalités essentielles : ré-indexation, suppression des entrées, sélection, filtrage, tri, agrégation
- Analyse descriptive : présentation et calcul des statistiques (corrélation, variance), l'appartenance, valeurs uniques
- Transformation des données en utilisant les fonctions

III. Arrays et calcul vectoriel

- Les bases de NumPy : création d'un vecteur, création d'un objet vecteur multidimensionnel, type de données pour ndarrays, transposition, tri, indexation

- Arithmétique avec NumPy arrays
- Exprimer la logique conditionnelle comme opération vectorielle

IV. **Chargement des données, stockage, et les formats des fichiers**

- Lire les données à partir des différents formats de fichiers (txt, csv, excel, json, ...)
- Ecrire les données à partir des différents formats de fichiers (csv, excel, json, ...)
- Interaction avec les bases de données

V. **Visualisation et outils d'affichage**

- Introduction à matplotlib : Figure et subplot, légende, labels
- Affichage avec les dataframes et seaborn (histogramme, bar et line plot, ..., etc.)

VI. **Manipulation des données textuelles**

- Méthodes d'objet String
- Expression régulière
- Fonction de string vectorisée dans pandas
- Représentation vectorielle des textes

VII. **Programmation orientée objet**

- Les classes et les objets
- Héritage et polymorphisme

VIII. **Packages et modules**

- Package hiérarchique de modules, l'import des submodules en utilisant des nom relative, chargement des modules, diviser un module en plusieurs fichiers
- Organisation du code et installation des packages

Mode d'évaluation : *Contrôle continu (TPs et mini projet), examen*

Références :

- Python Cookbook, Third Edition 3rd, Brian K. Jones, David M. Beazley, O'Reilly, 2011
- Fluent Python, Second Edition, Luciano Ramalho, O'Reilly, 2015
- Python by Example: Learning to Program in 150 Challenges, Nicola Lacey, Cambridge University, 2019
- Python for Data Analysis , Third Edition, Wes McKinney, O'Reilly, 2022
- Pandas for Everyone: Python Data Analysis, Second Edition, Daniel Y. Chen, Addison-Wesley Professional, 2017

Intitulé du Master : DATA SCIENCE and NATURAL LANGUAGE PROCESSING

Semestre : S1

Intitulé de l'UE : UEF1.2

Intitulé de la matière : Bases de données avancées

Crédits : 6

Coefficients : 3

Objectifs de l'enseignement (*Décrire ce que l'étudiant est censé avoir acquis comme compétences après le succès à cette matière – maximum 3 lignes*).

Maîtriser le concept « Base de Données », présenter les principaux modèles de données, concevoir une Base de Données relationnelle, et savoir créer et gérer une Base de Données sur MS Access.

Connaissances préalables recommandées : Notions d'informatique

Contenu de la matière : (*descriptif succinct des connaissances requises pour pouvoir suivre cet enseignement – Maximum 2 lignes*).

Cours 1 : Modèle Relationnel

- Partie 0: Rappel et introduction aux bases de données
- Partie 1: Modèle Entité association / étendu
- Partie 2: Modèle relationnel
- Partie 3: Passage Modèle E/A au modèle Relationnel)
- Partie 4: Le langage SQL

Cours 2 : Les SGBDs Déductifs

- Partie 0: La logique des prédicats
- Partie 1: Concepts de bases des bases de données déductives
- Partie 2: La réalisation d'un SGBD déductif
- Partie 3: Le langage DATALOG

Cours 3 : Bases de données Réparties

- Partie 0: Principe fondamental et Quelques termes équivoques
- Partie 1: Conception de BD répartie
- Partie 2: Schémas d'une BD répartie
- Partie 3: Fragmentation des données
- Partie 4: Requête répartie
- Partie 5: Gestion de l'hétérogénéité

Cours 4 : PL/SQL et Bases de données Actives

- Partie 1: Le langage PL/SQL
- Partie 2: Les déclencheurs

Cours 5: Bases de données et SGBD orientés Objet

- Partie 1: Introduction à la modélisation Orientée Objets
- Partie 2: SQL3 [Types , Tables OBJETS, Tables imbriquées , Méthodes ...etc.]

Méthodes d'évaluation :

- Examen écrit.
- Travaux pratiques en classe.

Références (*Livres et photocopiés, sites internet, etc*).

- Georges Gardarin. Bases de données: objet et relationnel. Eyrolles, 1999.
- Raghu Ramakrishnan, Johannes Gehrke. Database Management Systems. 2nd edition. Mc Graw-Hill, 1999.
- Tamer Özsu, Patrick Valduriez. Principles of Distributed Database Systems. 2nd edition, Prentice Hall, 1999.

Intitulé du Master : DATA SCIENCE and NATURAL LANGUAGE PROCESSING

Semestre : S1

Intitulé de l'UE : UEF1.3

Intitulé de la matière : Algorithmique Avancée et Calcul de complexité

Crédits : 6

Coefficients : 3

Objectifs de l'enseignement :

- Apprendre à effectuer une analyse formelle des algorithmes basée sur leurs complexités temporelle et spatiale.
- Évaluer l'efficacité des solutions (algorithmes de résolution) avant même de les implémenter, une étape nécessaire dans la résolution de n'importe quel problème.
- Classifier les problèmes selon leur degré de complexité.
- Utiliser des méthodes approchées pour résoudre les problèmes complexes.

Connaissances préalables recommandées :

- Algorithmique et structure de données.
- Théorie de la programmation.
- Théorie des graphes.

Contenu de la matière :

I. La complexité de calcul

- Exemples introductifs
- Algorithme et complexité
- Complexité théorique : Complexité temporelle vs complexité spatiale
- Complexité temporelle.
 - Définition et notion de landau
 - Règles de calcul
 - Classes de complexité

II. Structures de données élémentaires et avancées.

- Introduction aux structures de données
- Structures de données élémentaires Listes chaînées Les Piles Les files
- Structures de données avancées Les arbres binaires et les arbres binaires de recherche Les arbres binaires équilibrés: structure du TAS, AVL

III. Techniques de conception d'algorithmes

- La récursivité
- Technique 'Diviser et Conquérir'
- La programmation dynamique

IV. Algorithmes de tri.

- Les algorithmes classiques de tri

Tri Sélection, Tri par Insertion, Tri à Bulles

Les algorithmes de tri plus performants

Le tri Rapide, le tri par Tas, le tri Fusion

V. Les algorithmes de graphe.

- a. Rappels et notions élémentaires
- b. Parcours d'un arbre (graphe)
 - Parcours en profondeur d'abord
 - Parcours en largeur d'abord
- c. Recherche du plus court chemin
 - Algorithme de Dijkstra
- d. Arbre couvrant de poids minimal
 - Algorithme de Kruskal

VI. Introduction à la NP-complétude

- a. Introduction et notions fondamentales
- b. Problème, Instance de problème et solution
- c. Types de problèmes (Décision, Optimisation, Dénombrement de solutions,...)
- d. Classes de problèmes
 - La classe P
 - La classe NP
 - La classe NP-complet

Mode d'évaluation : Examens Ecrits (60%) + Contrôle continu (40%)

Références (*Livres et polycopiés, sites internet, etc*).

- Patrick Siarry, Johann Dréo, Alain Pétrowski, Éric Taillard : Métaheuristiques pour l'optimisation difficile, Eyrolles, 2003
- **Éléments d'algorithmique** Auteur(s) **Beauquier**, Danièle; **Berstel**, Jean; Chrétienne, Philippe. Edition masson 2003.
- Computers and Intractability: A Guide to the Theory of NP-Completeness (Michael R. Garey and David S. Johnson) 2006.

Intitulé du Master : DATA SCIENCE and NATURAL LANGUAGE PROCESSING

Semestre : S1

Intitulé de l'UE : UEM1.1

Intitulé de la matière : Computational Linguistics

Crédits : 3

Coefficients : 2

Objectifs de l'enseignement (*Décrire ce que l'étudiant est censé avoir acquis comme compétences après le succès à cette matière – maximum 3 lignes*).

aims to equip students with the skills to bridge linguistics and computer science, delving into the principles of natural language processing (NLP) and computational techniques. Students explore language representation, syntactic parsing, and semantic analysis, gaining a deep understanding of how computers can process and understand human language. The course also introduces tools and methodologies essential for tasks such as machine translation, speech processing, and corpus linguistics, fostering proficiency in applying computational approaches to linguistic analysis.

Connaissances préalables recommandées (*descriptif succinct des connaissances requises pour pouvoir suivre cet enseignement – Maximum 2 lignes*).

Students should have a basic knowledge of computer science and programming, as well as some experience in data processing. Additionally, familiarity with mathematics, statistics, and language theory would also be beneficial.

Contenu de la matière :

- Part 01: Natural Language Understanding
 - Chapter 01: Named Entity Recognition (NER)
 - Chapter 02: Part of Speech Tagging (PoS)
 - Chapter 04: Latent Dirichlet Allocation & Topics modelling
 - Chapter 05: Co-reference Resolution
 - Chapter 06: Representing text in vectors
 - Chapter 07: Deep learning methods for NLP
 - Chapter 08: Sequence labelling and classification.
- Part 02 : Natural Language Generation
 - Chapter 01: Natural Language modelling
 - Chapter 02: Lexical semantics
 - Chapter 03: Text generation
 - Chapter 04: Large Language Models and Introduction to Prompt engineering

Mode d'évaluation : Continu + Examen

Références (*Livres et photocopiés, sites internet, etc*).

- Jurafsky, D. & Martin, J.H. (2020). Speech and Language Processing (3rd ed.). Pearson.
- Manning, C.D. & Schütze, H. (1999). Foundations of Statistical Natural Language Processing. MIT Press.

- Bird, S., Klein, E. & Loper, E. (2009). Natural Language Processing with Python. O'Reilly Media.
- FUCHS Catherine, *Linguistique et traitements automatiques des langues*, Paris, Hachette, 1993.
- *Revue Maghrébine des Langues, Editions Dar El gharb (2000-2009), Volet Traductologie et TAL.*

Intitulé du Master : DATA SCIENCE and NATURAL LANGUAGE PROCESSING

Semestre : S1

Intitulé de l'UE : UEM1.2

Intitulé de la matière : Théorie des langages Formels

Crédits : 3

Coefficients : 2

Objectifs de l'enseignement (*Décrire ce que l'étudiant est censé avoir acquis comme compétences après le succès à cette matière – maximum 3 lignes*).

Devoir connaître et maîtriser les concepts issus de la théorie des langages et plus particulièrement des langages algébriques, de grammaire, de dérivateur, d'automate à pile qui sont en fait à la base de tous les algorithmes d'analyse syntaxiques.

Connaissances préalables recommandées (*descriptif succinct des connaissances requises pour pouvoir suivre cet enseignement – Maximum 2 lignes*).

Notions de mathématiques.

Contenu de la matière :

- Les langages
- Les automates d'états finis
- Les langages réguliers
- Les langages algébriques
- Les langages à contexte lié
- Les machines de Turing

Mode d'évaluation : *Contrôle continu, TP, examen final*

Références (*Livres et polycopiés, sites internet, etc.*).

- H.Hopcroft, D.Ullman.: Introduction to automata, theory languages and computation.
- M.Gross and A.Lentiu: Introduction to formal grammars.
- Patrice Séebold: Théorie des automates.
- A.V.Aho and J.D.Ullman: Principles of Compiler Design

Intitulé du Master : DATA SCIENCE and NATURAL LANGUAGE PROCESSING

Semestre : S1

Intitulé de l'UE : UEM1.3

Intitulé de la matière : Traitement des Données Linguistiques

Crédits : 3

Coefficients : 2

Objectifs de l'enseignement (*Décrire ce que l'étudiant est censé avoir acquis comme compétences après le succès à cette matière – maximum 3 lignes*).

Cette matière permet à l'étudiant de maîtriser les aspects articulatoires et acoustiques liés à la description phonétique.

Connaissances préalables recommandées (*descriptif succinct des connaissances requises pour pouvoir suivre cet enseignement – Maximum 2 lignes*).

Il s'agit d'une introduction au domaine, aucune compétence particulière n'est requise au départ si ce n'est un intérêt au domaine de la description scientifique de la langue.

Contenu de la matière :

1. Introduction à la phonétique générale
 - a. Phonétique acoustique
 - b. Phonétique perceptive
 - c. Phonétique articulatoire
2. Modes et lieu d'articulation
 - a. Consonnes
 - b. Voyelles
3. Phonétique arabe
4. Paramètres acoustiques
 - a. Fréquence fondamentale
 - b. Durée
 - c. Intensité
 - d. Formant
5. Paramètres suprasegmentaux
 - a. Mélodie
 - b. Rythme
6. Analyse de la parole sur un logiciel.

Mode d'évaluation : *Contrôle continu, examen final*

Références (*Livres et photocopiés, sites internet, etc*).

- *Livres de phonétique générale*

Intitulé du Master : DATA SCIENCE and NATURAL LANGUAGE PROCESSING

Semestre : S1

Intitulé de l'UE : UED1.1

Intitulé de la matière : Mathématiques Appliquées aux sciences de données

Crédits : 2

Coefficients : 1

Objectifs de l'enseignement (*Décrire ce que l'étudiant est censé avoir acquis comme compétences après le succès à cette matière*).

- Permettre à l'étudiant d'acquérir le savoir et la pratique des méthodes mathématiques appliquées pour la résolution des problèmes d'ingénierie.
- Maîtrise des méthodes mathématiques requises par les autres unités d'enseignements du master TRM

Connaissances préalables recommandées (*descriptif succinct des connaissances requises pour pouvoir suivre cet enseignement*).

Notions de mathématiques élémentaires

Contenu de la matière :

- Rappels des fondamentaux en Probabilités
- Fonctions de variables aléatoires
- Espérance mathématique et théorème de centrale limite
- Processus aléatoires et Ergodicité
- Processus markoviens
- Modèles bayésiens
- Rappels des fondamentaux en Statistiques
- Notions de base en Algèbre linéaire

Mode d'évaluation : *Contrôle continu, TP, examen final*

Références (*Livres et photocopiés, sites Internet, etc*).

- *Probabilités, analyse des données et statistique*, Gilbert Saporta, Editions technip
- *Mathématiques pour les sciences de l'ingénieur*, D. Fredon, M. Bridier, Lavoisier 2003
- Jaan Kiusallaas, *Numerical Methods in Engineering with Matlab*, Cambridge University Press Edt., 2005
- HWEI HSU, *Probability, Random variables and random processes in Schaum's Outlines* 1997
- Ionut Danaila and Co.
- *An Introduction to Scientific Computing : Twelve Computational Projects Solved with Matlab*, Springer Edt. 2007

Intitulé du Master : DATA SCIENCE and NATURAL LANGUAGE PROCESSING

Semestre : S1

Intitulé de l'UE : UET1.1

Intitulé de la matière : Sécurité des Systèmes Informatiques

Crédits : 1

Coefficients : 1

Objectifs de l'enseignement (*Décrire ce que l'étudiant est censé avoir acquis comme compétences après le succès à cette matière – maximum 3 lignes*).

Le module Sécurité des systèmes informatiques a pour objectif de fournir aux étudiants les connaissances de base en matière de sécurité informatique et de les initier aux techniques de protection des systèmes informatiques contre les attaques. Les objectifs spécifiques comprennent la compréhension des menaces courantes, la mise en place de stratégies de sécurité, l'utilisation d'outils de sécurité et la compréhension des enjeux liés à la sécurité.

Connaissances préalables recommandées (*descriptif succinct des connaissances requises pour pouvoir suivre cet enseignement – Maximum 2 lignes*).

Les étudiants devraient avoir une connaissance de base en informatique, en programmation et en réseau. Des connaissances en mathématiques et en cryptographie seraient également utiles.

Contenu de la matière :

- Introduction à la sécurité informatique
- Menaces courantes : virus, vers, chevaux de Troie, phishing, déni de service, etc.
- Techniques de protection : pare-feu, antivirus, chiffrement, signatures numériques, authentification, etc.
- Stratégies de sécurité : politique de sécurité, gestion des mots de passe, sécurité des réseaux, sécurité des applications, etc.
- Enjeux liés à la sécurité : confidentialité, intégrité, disponibilité, conformité, etc.

Mode d'évaluation : *Contrôle continu, examen final*

Références (*Livres et polycopiés, sites internet, etc*).

- Stallings, W. (2017). Network Security Essentials: Applications and Standards (6th ed.). Pearson.
- Pfleeger, C. P., & Pfleeger, S. L. (2018). Security in Computing (5th ed.). Pearson.
- Ross, G. (2015). Introduction to Information Security: A Strategic-Based Approach. Wiley.

Intitulé du Master : DATA SCIENCE and NATURAL LANGUAGE PROCESSING

Semestre : S2

Intitulé de l'UE : UEF1.1

Intitulé de la matière : Apprentissage Automatique et Applications Neuronales

Crédits : 6

Coefficients : 3

Objectifs de l'enseignement

Ce cours a pour objectif essentiel de présenter la théorie et les outils de la RdF qui se rattachent à ces différents aspects dans le cadre des applications faisant appel à l'apprentissage automatique.

Les TP auront pour but de tester différents algorithmes sur des problèmes simples 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

- **Introduction** : Concepts de base et terminologie, Exemples d'application de la RDF, Visualisation et prétraitement des données, calcul des représentations, Fondamentaux du machine learning.
- **Méthodes d'apprentissage supervisée** :
 - Les méthodes de discrimination paramétriques : Naive bayes
 - Les méthodes de discrimination non paramétriques : k-plus proches voisins.
 - Les machines à vecteurs de supports
 - Arbres de décision
 - Evaluation d'un algorithme : échantillon de test, validation croisée, courbe ROC
- **Techniques de regroupement (Clustering)** : K-means, classification hiérarchique ascendante, méthodes basées densité, méthodes basées grille, Expectation-Maximisation.
- **Techniques neuronales** :
 - Le Perceptron : Application à la classification et à la régression
 - Le Perceptron multicouches : Application à la classification et à la régression
 - Les cartes auto-organisatrices
 - Introduction aux réseaux de neurones convolutifs et à l'apprentissage profond.
- **Réduction de la dimensionnalité de données** :
 - Sélection d'attributs : en avant, en arrière
 - Extraction d'attributs: Analyse à composantes principales (ACP) ; Analyse discriminante linéaire (ADL)

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 : DATA SCIENCE and NATURAL LANGUAGE PROCESSING

Semestre : S2

Intitulé de l'UE : UEF1.2

Intitulé de la matière : Data Mining

Crédits : 6

Coefficients : 3

Objectifs de l'enseignement :

- Apprendre à connaître les données.
- Connaître et savoir appliquer les fondamentaux théoriques et pratiques de la fouille de données.
- Connaître et savoir appliquer les techniques d'extraction de connaissance à partir de données quelconques.

Connaissances préalables recommandées :

- Mathématique.
- Algorithmique et structure de données.

Contenu de la matière :

I. Introduction à la fouille de données

II. Etude statistique de la donnée

- a. Type de données
- b. Etude statistique des données
- c. Calcul de similarité selon le type de données

III. Prétraitement des données (nettoyage, normalisation, réduction des données)

- a. Nettoyage des données (données aberrantes, manquantes).
- b. Intégration et normalisation des données
- c. Réduction des données

IV. Règles d'association, motifs fréquents et corrélation

- a. Notions fondamentales
- b. Motifs fréquents et règles d'associations
- c. Algorithme Apriori

V. Clustering

- a. Introduction au clustering
- b. Clustering selon type de données
- c. Méthodes de clustering
 1. Par partitionnement (K-medoid et Clarans)
 2. Hiérarchique
 3. Selon densité (DBSCAN)
 4. Selon grille
- d. Outlier detection

VI. Classification et Prédiction

- a. Introduction, classification vs prediction
- b. Méthodes de classification

Choix des attributs et des méthodes Arbres de décision et Réseaux bayésiens

- c. Evaluation d'un classifieur (métriques)

Mode d'évaluation : Examens Ecrits (60%) + Contrôle continu (40%)

Références (*Livres et photocopiés, sites internet, etc*).

- Carl Shan, William Chen, Henry Wang et Max Song, The Data Science Handbook: Advice and Insights from 25 Amazing Data Scientists, 2015.
- Roger D. Peng et Elizabeth Matsui, The Art of Data Science, 2016

Intitulé du Master : DATA SCIENCE and NATURAL LANGUAGE PROCESSING

Semestre : S2

Intitulé de l'UE : UEF1.3

Intitulé de la matière : Bases de Données NoSQL pour le Big Data

Crédits : 6

Coefficients : 3

Objectifs de l'enseignement (*Décrire ce que l'étudiant est censé avoir acquis comme compétences après le succès à cette matière – maximum 3 lignes*).

Ce cours permettra aux étudiants de comprendre les principes fondamentaux de la base de données NoSQL pour le Big Data. Les étudiants exploreront les différentes technologies NoSQL, telles que HBase, Cassandra et MongoDB, ainsi que les applications Big Data telles que Hadoop, MapReduce. Ils apprendront également comment choisir la technologie NoSQL appropriée pour leur application Big Data.

Objectifs d'apprentissage :

- Comprendre les concepts de base des bases de données NoSQL et leur évolution
- Comprendre les différents types de bases de données NoSQL, tels que les bases de données orientées colonnes, les bases de données orientées documents, les bases de données clé-valeur et les bases de données graphiques
- Comprendre les caractéristiques de performance et les cas d'utilisation de chaque type de base de données NoSQL
- Comprendre les outils de Big Data, tels que Hadoop, Spark et MapReduce
- Comprendre comment les bases de données NoSQL sont utilisées dans les environnements Big Data
- Comprendre comment sélectionner la technologie NoSQL appropriée pour une application Big Data donnée
- Comprendre comment concevoir et mettre en œuvre des systèmes de base de données NoSQL pour le Big Data

Contenu du cours :

- **Introduction aux bases de données NoSQL**
 - o Définition et objectifs des bases de données NoSQL
 - o Différences avec les bases de données relationnelles
 - o Types de bases de données NoSQL (document, graph, clé-valeur, colonne, etc.)
- **Modélisation de données pour NoSQL**
 - o Concepts de modélisation de données NoSQL
 - o Schémas dynamiques et flexibles
- **NoSQL et le Big Data**
 - o Caractéristiques et défis du Big Data
 - o Bases de données NoSQL pour le Big Data
 - o Utilisation de NoSQL pour le stockage et l'analyse de données massives [Contexte de données textuelles]

Les bases de données NoSQL les plus populaires

- MongoDB
- Cassandra
- Neo4j
- DynamoDB
- HBase
- Couchbase
- **Les opérations de base sur les bases de données NoSQL**
 - Lecture et écriture de données
 - Requêtes et agrégations
 - Indexation et optimisation des performances
- **Cas d'utilisation de NoSQL pour le Big Data**
 - Exemples de cas d'utilisation : Traitement des données de réseaux sociaux, Traitement de bases de données documentaires, etc.

Méthodes d'évaluation :

- Examen écrit.
- Travaux pratiques en classe.
- Projet de conception et de mise en place d'une base de données NoSQL pour le Big Data[Application aux TAL].

Intitulé du Master : DATA SCIENCE and NATURAL LANGUAGE PROCESSING

Semestre : S2

Intitulé de l'UE : UEM1.1

Intitulé de la matière : Principes de Conception et Inférence Causale pour la Prise de Décision Basée sur les Données

Crédits : 3

Coefficients : 2

Objectifs de l'enseignement :

Cette matière vise à fournir aux étudiants en data science les compétences nécessaires pour concevoir des modèles de données robustes et effectuer des inférences causales. Les objectifs incluent la compréhension des principes fondamentaux de la causalité, la maîtrise des méthodes de conception de modèles causaux, et l'application de ces concepts à la prise de décision basée sur les données.

Connaissances préalables recommandées :

Les étudiants doivent avoir une solide compréhension des statistiques, de l'algèbre linéaire, et des compétences de programmation en Python ou R. Une familiarité avec les concepts de base de la data science et de l'apprentissage automatique serait également bénéfique.

Contenu de la matière :

- Introduction à la Causalité
 - Définitions et concepts de base en causalité.
 - Différence entre corrélation et causalité.
- Conception de Modèles Causaux
 - Graphes de causalité et notation.
 - Identification des relations causales et des variables confondantes.
- Inférence Causale
 - Méthodes d'estimation des effets causaux.
 - Tests de robustesse et sensibilité des résultats.
- Applications à la Prise de Décision
 - Utilisation de modèles causaux pour la prise de décision.
 - Intégration des résultats dans un contexte décisionnel.

Mode d'évaluation : Examens Ecrits (60%) + Contrôle continu (40%)

Références (Livres et polycopiés, sites internet, etc).

- "Causal Inference in Statistics: A Primer" par Judea Pearl (2009).
- "Counterfactuals and Causal Inference: Methods and Principles for Social Research" par Stephen L. Morgan et Christopher Winship (2014).
- "Causality: Models, Reasoning, and Inference" par Judea Pearl (2009).
- "The Book of Why: The New Science of Cause and Effect" par Judea Pearl et Dana Mackenzie (2018).
- "Causal Inference: What If" par Miguel Hernan et James M. Robins (2020).

Intitulé du Master : DATA SCIENCE and NATURAL LANGUAGE PROCESSING

Semestre : S2

Intitulé de l'UE : UEM1.2

Intitulé de la matière : Entrepôts de données et Analyse OLAP

Crédits : 3

Coefficients : 2

Objectifs de l'enseignement (*Décrire ce que l'étudiant est censé avoir acquis comme compétences après le succès à cette matière – maximum 3 lignes*).

Apprendre la conception d'un entrepôt de données, le processus ETL ainsi que la maîtrise des opérateurs OLAP. Introduire la technologie de stockage des données volumineuse : Les lacs de données (Data Lake).

Connaissances préalables recommandées (*descriptif succinct des connaissances requises pour pouvoir suivre cet enseignement – Maximum 2 lignes*).

Contenu de la matière (*indiquer obligatoirement le contenu détaillé du programme en présentiel et du travail personnel*)

- Généralités sur les systèmes d'informations décisionnelles.
- Modélisation d'un entrepôt de données.
- Processus ETL : Extraction, Transformation, Chargement.
- Analyse en ligne de données (OLAP) et opérateurs d'analyse.
- Lac de données pour l'analyse des Big Data.

Mode d'évaluation : *Contrôle continu, examen, etc... (La pondération est laissée à l'appréciation de l'équipe de formation)*

Références (*Livres et photocopiés, sites internet, etc*).

- E. F. Codd, S. B. Codd, and C. T. Salley. Providing olap to user-analysts : An it mandate. Ann Arbor Michigan, 26 :24, 1993.
- W. H. Inmon. Building the data warehouse (2nd ed.). John Wiley & Sons, Inc., New York, NY, USA, 1996.
- Kimball. The Data Warehouse Toolkit : Practical Techniques for Building Dimensional Data Warehouses. John Wiley, 1996.

Intitulé du Master : DATA SCIENCE and NATURAL LANGUAGE PROCESSING

Semestre : S2

Intitulé de l'UE : UEM1.3

Intitulé de la matière : Algorithmique Avancée et Résolution de Problèmes complexes

Crédits : 3

Coefficients : 2

Objectifs de l'enseignement :

- Différencier les techniques de résolution classiques et intelligentes.
- Avoir les capacités de raisonner devant un problème complexe.
- Construire une solution dans le domaine de la résolution de problèmes.
- Evaluer l'efficacité d'une solution.
- Appliquer les différentes techniques acquises sur des problèmes réels.

Connaissances préalables recommandées :

- Algorithmique avancé et complexité de calcul.
- Théorie de la programmation.
- Théorie des graphes.

Contenu de la matière :

I. Les méthodes exactes pour la résolution des problèmes complexes.

- a. Exemples de problèmes combinatoires
- b. Notions fondamentales à la résolution de problèmes
- c. Méthodes de recherche aveugles
- d. Méthodes heuristiques

II. Les méthodes approchées et algorithmes évolutionnaires pour la résolution des problèmes complexes.

- a. Algorithmes évolutionnaires
- b. Méta-heuristique et Intelligence en essaim
Introduction et type de méta-heuristique
 1. Recherche locale et recherche taboue
 2. Les algorithmes génétiques
 3. Les essaims d'abeilles
 4. Les colonies de fourmisL'hybridation entre méta-heuristique

Mode d'évaluation : Examens Ecrits (60%) + Contrôle continu (40%)

Références (Livres et polycopiés, sites internet, etc).

- Patrick Siarry, Johann Dréo, Alain Pétrowski, Éric Taillard : Métaheuristiques pour l'optimisation difficile, Eyrolles, 2003
- **Eléments d'algorithmique** Auteur(s) **Beauquier**, Danièle; **Berstel**, Jean; Chrétienne, Philippe. Edition masson 2003.
- Computers and Intractability: A Guide to the Theory of NP-Completeness (Michael R. Garey and David S. Johnson) 2006

Intitulé du Master : DATA SCIENCE and NATURAL LANGUAGE PROCESSING

Semestre : S2

Intitulé de l'UE : UET1.1

Intitulé de la matière : Business Intelligence

Crédits : 1

Coefficients : 1

Objectifs de l'enseignement (*Décrire ce que l'étudiant est censé avoir acquis comme compétences après le succès à cette matière*).

Le module Business Intelligence (BI) a pour objectif de fournir aux étudiants les compétences nécessaires pour collecter, analyser et visualiser des données d'entreprise pour prendre des décisions éclairées. Les objectifs spécifiques comprennent la compréhension des concepts et des outils de BI, l'utilisation de techniques d'analyse de données pour la prise de décisions, la conception de tableaux de bord pour la présentation de données, et la mise en place de systèmes de BI pour les entreprises.

Connaissances préalables recommandées (*descriptif succinct des connaissances requises pour pouvoir suivre cet enseignement*).

Les étudiants devraient avoir une connaissance de base en informatique, en mathématiques et en statistiques. Des compétences en analyse de données seraient également utiles.

Contenu de la matière :

Le module couvre les sujets suivants :

- *Introduction à la Business Intelligence*
- *Collecte de données : ETL, intégration de données*
- *Analyse de données : requêtes, agrégation, exploration de données, data mining*
- *Visualisation de données : tableaux de bord, graphiques, etc.*
- *Implémentation de systèmes de BI : entrepôts de données, architecture de BI, sécurité*

Mode d'évaluation : *Contrôle continu, examen final*

Références (*Livres et polycopiés, sites Internet, etc*).

- *Kimball, R., Ross, M., Thornthwaite, W., Mundy, J., & Becker, B. (2013). The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling (3rd ed.). Wiley.*
- *Microsoft. (2021). Power BI Documentation. <https://docs.microsoft.com/en-us/power-bi/>*
- *Shmueli, G., Patel, N. R., & Bruce, P. C. (2020). Data Mining for Business Analytics: Concepts, Techniques, and Applications in Python. Wiley.*

Intitulé du Master : DATA SCIENCE and NATURAL LANGUAGE PROCESSING

Semestre : S3

Intitulé de l'UE : UEF1.1

Intitulé de la matière : Big data Analytics

Crédits : 6

Coefficients : 3

Objectifs de l'enseignement : Ce cours a pour but d'initier les étudiants à la résolution de problèmes en linguistique informatique.

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 : Algèbre linéaire / concepts de base en apprentissage automatique

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 realtime 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 : DATA SCIENCE and NATURAL LANGUAGE PROCESSING

Semestre : S3

Intitulé de l'UE : UEF1.2

Intitulé de la matière : Business Intelligence

Crédits : 6

Coefficients : 3

Objectifs de l'enseignement : Ce cours a pour but d'initier les étudiants à la résolution de problèmes en linguistique informatique.

Connaissances préalables recommandées : Algèbre linéaire / concepts de base en apprentissage automatique

Contenu de la matière :

- Introduction à La modélisation linguistique :
 - o Introduction aux modèles de langage
 - o Les modèles n-grammes
 - o Lacunes des modèles classiques
- La représentation neuronale linguistique
 - o Introduction aux réseaux de neurones
 - o La représentation en vecteurs de mots (ou word embeddings) :
 - CBOW , SKIP-GRAM , GLOVE
- Les réseaux de neurones récurrents (RNN) :
 - o Définition du modèle linguistique récurrents
 - o Problèmes de vanishing/exploiding gradient
 - o Introduction aux LSTM/GRU
 - o Les RNN bidirectionnels
 - o Les modèles sequence-to-sequence
- Les Transformers pour la modélisation linguistique
 - o Définition des modèles encoder-decoder
 - o Les mécanismes d'attention : attention (globale, locale) / self-attention
 - o Modèles linguistiques à base de transformers :
 - BERT
 - GPT
- **Applications de l'algorithmique et de la modélisation linguistique:** Classification de documents, Recherche d'information, analyse de sentiments, traduction automatique, génération automatique de résumé,

Mode d'évaluation : $(2 \times \text{note examen final} + \text{note contrôle continu}) / 3$

Références (Livres et photocopiés, sites internet, etc).

- "Neural Network Methods in Natural Language Processing», Yoav Goldberg, Springer, 2017
- "Deep Learning for Natural Language Processing", Stephan Raaijmakers, Manning, 2018

Intitulé du Master : DATA SCIENCE and NATURAL LANGUAGE PROCESSING

Semestre : S3

Intitulé de l'UE : UEF1.3

Intitulé de la matière : Méthodes chronologiques et prévisions

Crédits : 6

Coefficients : 3

Objectifs de l'enseignement (Décrire ce que l'étudiant est censé avoir acquis comme compétences après le succès à cette matière – maximum 3 lignes).

Ce module vise à doter les étudiants en data science des compétences nécessaires pour développer des applications web interactives permettant la visualisation et le déploiement de résultats issus de l'analyse de données. Les objectifs incluent la compréhension des technologies web, la création d'interfaces utilisateur intuitives, et l'intégration de fonctionnalités interactives pour présenter des conclusions data-driven.

Connaissances préalables recommandées (descriptif succinct des connaissances requises pour pouvoir suivre cet enseignement – Maximum 2 lignes).

Les étudiants doivent avoir une connaissance de base en programmation (idéalement en Python ou JavaScript) et une compréhension des concepts fondamentaux de la data science. Une familiarité avec les bases de données et la manipulation de données est également recommandée.

Contenu de la matière :

- Introduction
- Développement front-end et back-end.
- Le Langage HTML et les feuilles de styles.
- Création des effets simples avec les feuilles de styles (effets de transitions, menus déroulants, surbrillance du texte, effets d'animation).
- Programmation côté client (Javascript, jQuery et Ajax).
- Conception et développement de sites web responsif dont le contenu et la mise en page s'adaptent aux périphériques utilisés (écrans d'ordinateur, télévisions, téléphones, tablettes, etc.).
- Assurer la compatibilité d'un site avec différents navigateurs.
- Les animations simples.

Mode d'évaluation : Contrôle continu (40%) , examen (60%)

Références bibliographiques (Livres et polycopiés, sites internet, etc) :

Citer au moins 3 à 4 références classiques et importantes.

- "Flask Web Development" par Miguel Grinberg (2018).
- "JavaScript: The Good Parts" par Douglas Crockford (2008).
- "D3.js in Action" par Elijah Meeks (2015).
- "Designing Data-Intensive Applications" par Martin Kleppmann (2017).
- Documentation officielle des frameworks et bibliothèques utilisés (Flask, D3.js, etc.).

Intitulé du Master : DATA SCIENCE and NATURAL LANGUAGE PROCESSING

Semestre : S3

Intitulé de l'UE : UEM1.1

Intitulé de la matière : Recherche d'Information

Crédits : 3

Coefficients : 2

Objectifs de l'enseignement (*Décrire ce que l'étudiant est censé avoir acquis comme compétences après le succès à cette matière – maximum 3 lignes*).

Exposer les fondements, méthodes et améliorations des systèmes de recherche d'information. Maîtriser la conception et l'implémentation d'un système de recherche d'information

Connaissances préalables recommandées (*descriptif succinct des connaissances requises pour pouvoir suivre cet enseignement – Maximum 2 lignes*).

- Programmation (*python*)
- Notions mathématiques et informatique

Contenu de la matière (*indiquer obligatoirement le contenu détaillé du programme en présentiel et du travail personnel*)

- Introduction et notions de bases de la recherche d'information
- Processus d'indexation automatique
- Techniques de pondération des termes
- Modèles de base de la recherche d'information : Booléen classique, vectoriel et booléen étendu.
- Le modèle LSI (*Latent Symantic Indexing*)
- Les modèles probabilistes et de langues
- Enrichissement de requêtes
- Recherche d'information par apprentissage (*Learning to Rank*)
- Evaluation des systèmes de recherche d'information

Mode d'évaluation : *Contrôle continu, examen, etc...* (*La pondération est laissée à l'appréciation de l'équipe de formation*)

Contrôle continu et examen final

Références (*Livres et photocopiés, sites internet, etc*).

- Van Rijsbergen., Information Retrieval, 2ème Edition, Butterworths, Londres (UK), 1979.
- Yates R.B., Neto R., Modern information retrieval. ACM Press, Addison Wesley, 1999.
- Salton G., The Smart Retrieval System : Experiments in Automatic Document Processing, - G. Salton Editor, Prentice Hall Inc., Englewood Cliffs, New Jersey, 1971.

- Salton G. Automatic Text Processing. The Transformation Analysis and Retrieval of Information by Computer. Addison Wesley, Reading 1989 - Zadeh L.A. Fuzzy Sets, Information Control, 8 : p 338-353, 1965
- Hofman T. Probabilistic Latent Semantic Indexing : In the Proceedings of the 22nd Annual - International ACM SIGIR, Conference on Research and Development in Information Retrieval, August, 1999, Buckley USA
- Robertson S.E., Sparck Jones K. Relevance Weighting for Search Terms, Journal of The American Society for Information Science, Vol 27, N°3, pp 129-146, 1976
- Azad, Hiteshwar Kumar, and Akshay Deepak. "Query expansion techniques for information retrieval: a survey." *Information Processing & Management* 56.5,1698-1735, 2019.
- Liu, Tie-Yan. "Learning to rank for information retrieval." *Foundations and Trends® in Information Retrieval* 3.3, 225-331, 2009.
- Li, Hang. "A short introduction to learning to rank." *IEICE TRANSACTIONS on Information and Systems* 94.10,1854-1862, 2011.
- Mitra, Bhaskar, and Nick Craswell. "An introduction to neural information retrieval." *Foundations and Trends® in Information Retrieval* 13.1,1-126, 2018.

Intitulé du Master : DATA SCIENCE and NATURAL LANGUAGE PROCESSING

Semestre : S3

Intitulé de l'UE : UEM1.2

Intitulé de la matière : Graphes de connaissances et apprentissage

Crédits : 3

Coefficients : 2

Objectifs de l'enseignement (*Décrire ce que l'étudiant est censé avoir acquis comme compétences après le succès à cette matière – maximum 3 lignes*).

Ce module vise à familiariser les étudiants avec les concepts avancés de représentation des connaissances à l'aide de graphes et à explorer les méthodes d'apprentissage sur ces structures. Les objectifs incluent la compréhension des modèles de graphes de connaissances, la mise en œuvre d'algorithmes d'apprentissage sur des données structurées en graphes, et l'application de ces compétences dans des contextes réels.

Connaissances préalables recommandées (*descriptif succinct des connaissances requises pour pouvoir suivre cet enseignement – Maximum 2 lignes*).

Les étudiants doivent avoir des connaissances de base en apprentissage automatique, en particulier dans les domaines de l'apprentissage supervisé et non supervisé. Une compréhension des structures de données, y compris les graphes, serait bénéfique.

Contenu de la matière :

1. Introduction aux Graphes de Connaissances
 - Définition des graphes de connaissances.
 - Utilisation de standards comme RDF (Resource Description Framework) et OWL (Web Ontology Language).
2. Modèles de Représentation des Connaissances
 - Représentation des entités et des relations dans un graphe.
 - Utilisation de schémas pour définir des ontologies.
3. Apprentissage sur les Graphes
 - Méthodes d'apprentissage supervisé et non supervisé sur des données en format de graphe.
 - Exploration des approches de propagation de l'information et de marche aléatoire.
4. Intégration de Données en Graphe
 - Extraction et intégration de données pour la construction de graphes de connaissances.
 - Gestion de l'hétérogénéité des sources de données.
5. Applications Pratiques
 - Exploration de cas d'utilisation concrets dans divers domaines (par exemple, recommandation, recherche sémantique).
 - Études de cas sur l'utilisation de graphes de connaissances dans l'industrie.
6. Représentation d'Entités et Incorporation de Graphes
 - Techniques d'embeddings pour représenter les entités dans des espaces vectoriels.
 - Utilisation d'incorporations de graphes pour capturer les relations entre les entités.

7. Évaluation et Optimisation

- Méthodes d'évaluation des modèles de graphes de connaissances.
- Optimisation des hyperparamètres et ajustement du modèle.

Quelques Références :

- "Knowledge Graphs" par Sujit Pal (2020).
- "Graph-Based Semi-Supervised Learning" par Xiaojin Zhu (2005).
- "Representation Learning on Graphs" par William L. Hamilton, Rex Ying, et Jure Leskovec (2017).
- "Semantic Web for the Working Ontologist" par Dean Allemang et Jim Hendler (2011).
- "Knowledge Graphs and Language Models for Data Science" par Michael Nicosia et Amit P. Sheth (2021).

Intitulé du Master : DATA SCIENCE and NATURAL LANGUAGE PROCESSING**Semestre : S3****Intitulé de l'UE : UEM1.3****Intitulé de la matière : Extraction, Analyse et visualisation des données****Crédits : 3****Coefficients : 2**

Objectifs de l'enseignement (*Décrire ce que l'étudiant est censé avoir acquis comme compétences après le succès à cette matière – maximum 3 lignes*).

Cette matière a pour objectif de fournir aux étudiants en data science les compétences nécessaires pour collecter, nettoyer, analyser et visualiser efficacement des données. Les objectifs incluent la compréhension des méthodes d'extraction de données, la maîtrise des techniques d'analyse exploratoire, et la création de visualisations informatives pour faciliter la compréhension des données.

Connaissances préalables recommandées (*descriptif succinct des connaissances requises pour pouvoir suivre cet enseignement – Maximum 2 lignes*).

Les étudiants doivent avoir une connaissance de base en statistiques, des compétences en programmation (de préférence en Python ou R), et une compréhension des concepts fondamentaux de la data science. Une familiarité avec les bases de données et les outils d'analyse de données serait également bénéfique.

Contenu de la matière :

- **Introduction à l'Extraction de Données**
 - Techniques d'extraction de données à partir de différentes sources.
 - Nettoyage initial des données : gestion des valeurs manquantes, détection des outliers.
- **Analyse Exploratoire des Données (AED)**
 - Méthodes d'AED pour comprendre la distribution des données.
 - Analyse des tendances, des corrélations et des motifs.
- **Outils et Technologies pour l'Analyse de Données**
 - Utilisation de bibliothèques Python (pandas, NumPy) ou R pour l'analyse de données.
 - Introduction à l'utilisation de SQL pour l'analyse de données dans les bases de données.
- **Visualisation de Données**
 - Principes de conception de visualisations efficaces.
 - Utilisation d'outils tels que Matplotlib, Seaborn, ggplot2, ou des outils de BI (Business Intelligence).
- **Visualisation Interactive et Tableaux de Bord**
 - Création de visualisations interactives avec des outils tels que Plotly ou Tableau.
 - Construction de tableaux de bord pour une présentation visuelle complète des résultats.

Quelques Références :

"Python for Data Analysis" par Wes McKinney (2017).
 "R for Data Science" par Hadley Wickham et Garrett Grolemund (2017).
 "Data Science for Business" par Foster Provost et Tom Fawcett (2013).
 "Storytelling with Data" par Cole Nussbaumer Knafllic (2015).
 "Interactive Data Visualization for the Web" par Scott Murray (2017).

Intitulé du Master : DATA SCIENCE and NATURAL LANGUAGE PROCESSING

Semestre : S3

Intitulé de l'UE : UED1.1

Intitulé de la matière : Entrepreneuriat

Crédits : 2

Coefficients : 1

Objectifs de l'enseignement (Décrire ce que l'étudiant est censé avoir acquis comme compétences après le succès à cette matière – maximum 3 lignes).

Ce module a pour but de fournir aux étudiants en data science les compétences nécessaires pour comprendre et appliquer des concepts entrepreneuriaux dans le contexte de la data science. Les objectifs incluent l'identification d'opportunités commerciales, la création de produits basés sur des analyses de données, et la compréhension des aspects pratiques de la création et de la gestion d'une entreprise en data science.

Connaissances préalables recommandées (descriptif succinct des connaissances requises pour pouvoir suivre cet enseignement – Maximum 2 lignes).

Aucun prérequis spécifique n'est nécessaire, mais une connaissance de base en data science et une passion pour l'innovation seront bénéfiques.

Contenu du Module :

- Introduction à l'Entrepreneuriat en Data Science
 - Concepts fondamentaux de l'entrepreneuriat.
 - Importance de l'innovation basée sur les données dans le secteur des start-ups.
- Identification d'Opportunités et Analyse du Marché
 - Méthodes pour identifier des opportunités commerciales dans le domaine de la data science.
 - Analyse du marché et évaluation de la demande pour les produits data-driven.
- Élaboration de Plans d'Affaires
 - Structuration d'un plan d'affaires pour une entreprise en data science.
 - Intégration des aspects liés à la data dans le modèle commercial.
- Développement de Produits Data-Driven
 - Processus de développement de produits basés sur des analyses de données.
 - Intégration de l'apprentissage automatique et de l'analyse prédictive dans le développement.
- Stratégies de Marketing et de Vente
 - Élaboration de stratégies de marketing pour les produits data-driven.
 - Techniques de vente spécifiques aux services et produits en data science.
- Financement et Levée de Fonds
 - Options de financement pour les start-ups en data science.
 - Préparation d'une levée de fonds et négociation avec des investisseurs.
- Gestion Opérationnelle et Légale
 - Gestion opérationnelle quotidienne d'une entreprise en data science.
 - Conformité légale et réglementaire dans le domaine de la data.
- Études de Cas et Retours d'Expérience
 - Analyse d'études de cas d'entreprises en data science réussies.
 - Retours d'expérience d'entrepreneurs du secteur.

Quelques Références :

- "The Lean Startup" par Eric Ries (2011).
- "Zero to One" par Peter Thiel (2014).

- *"Data-Driven: Creating a Data Culture" par Hilary Mason et DJ Patil (2015).*
- *"Crossing the Chasm" par Geoffrey A. Moore (1991).*
- *"Venture Deals" par Brad Feld et Jason Mendelson (2016).*

Intitulé du Master : DATA SCIENCE and NATURAL LANGUAGE PROCESSING

Semestre : S3

Intitulé de l'UE : UET1.1

Intitulé de la matière : Techniques de rédaction

Crédits : 1

Coefficients : 1

Objectifs de l'enseignement (*Décrire ce que l'étudiant est censé avoir acquis comme compétences après le succès à cette matière – maximum 3 lignes*).

- Acquérir les compétences nécessaires pour effectuer une recherche documentaire efficace dans le cadre d'un mémoire de Master
- Apprendre à formuler une question de recherche et à définir un sujet de mémoire de manière claire et précise
- Savoir comment sélectionner et évaluer des sources fiables et pertinentes pour la recherche
- Maîtriser les techniques de citation et de référencement pour éviter le plagiat
- Apprendre à rédiger un mémoire de Master clair, concis et bien structuré

Connaissances préalables recommandées (*descriptif succinct des connaissances requises pour pouvoir suivre cet enseignement – Maximum 2 lignes*).

- Avoir une bonne maîtrise de la langue utilisée pour rédiger le mémoire
- Connaissance de base de la méthodologie de la recherche scientifique
- Capacité à utiliser les outils informatiques pour effectuer des recherches documentaires

Contenu de la matière :

- Introduction à la recherche documentaire pour les mémoires de Master
- Formulation de la question de recherche et définition du sujet de mémoire
- Recherche d'information pertinente et fiable pour le mémoire
- Techniques de citation et de référencement pour éviter le plagiat
- Rédaction du mémoire de Master et structuration du contenu
- Préparation de la présentation orale du mémoire de Master

Mode d'évaluation: Continu + Examen

Références (*Livres et photocopiés, sites Internet, etc.*).

- Creswell, J. W. (2014). Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. Sage publications.
- Day, R. A., & Gastel, B. (2012). How to write and publish a scientific paper. Greenwood Publishing Group.
- Hart, C. (2018). Doing a Literature Review: Releasing the Research Imagination. Sage publications.
- Swales, J. M., & Feak, C. B. (2012). Academic Writing for Graduate Students: Essential Tasks and Skills. University of Michigan Press.

V. Accords ou conventions

LETTRE D'INTENTION TYPE

(En cas de master coparrainé par un autre établissement universitaire)

(Papier officiel à l'entête de l'établissement universitaire concerné)

Objet : Approbation du coparrainage du master intitulé :

Par la présente, l'université (ou le centre universitaire) déclare coparrainer le master ci-dessus mentionné durant toute la période d'habilitation de ce master.

A cet effet, l'université (ou le centre universitaire) assistera ce projet en :

- Donnant son point de vue dans l'élaboration et à la mise à jour des programmes d'enseignement,
- Participant à des séminaires organisés à cet effet,
- En participant aux jurys de soutenance,
- En œuvrant à la mutualisation des moyens humains et matériels.

SIGNATURE de la personne légalement autorisée :

FONCTION :

Date :

LETTRE D'INTENTION TYPE

(En cas de master en collaboration avec une entreprise du secteur utilisateur)

(Papier officiel à l'entête de l'entreprise)

OBJET : Approbation du projet de lancement d'une formation de master intitulé :

Dispensé à :

Par la présente, l'entreprise déclare sa volonté de manifester son accompagnement à cette formation en qualité d'utilisateur potentiel du produit.

A cet effet, nous confirmons notre adhésion à ce projet et notre rôle consistera à :

- Donner notre point de vue dans l'élaboration et à la mise à jour des programmes d'enseignement,
- Participer à des séminaires organisés à cet effet,
- Participer aux jurys de soutenance,
- Faciliter autant que possible l'accueil de stagiaires soit dans le cadre de mémoires de fin d'études, soit dans le cadre de projets tuteurés.

Les moyens nécessaires à l'exécution des tâches qui nous incombent pour la réalisation de ces objectifs seront mis en œuvre sur le plan matériel et humain.

Monsieur (ou Madame).....est désigné(e) comme coordonateur externe de ce projet.

SIGNATURE de la personne légalement autorisée :

FONCTION :

Date :

CACHET OFFICIEL ou SCEAU DE L'ENTREPRISE

VI. Curriculum Vitae des Coordinateurs

VII - Avis et Visas des organes administratifs et consultatifs

Intitulé du Master : DATA SCIENCE and NATURAL LANGUAGE PROCESSING

Comité Scientifique de département	
Avis et visa du Comité Scientifique : <i>Favorable</i>	
Date : <i>25/02/2024</i>	<i>محمد ولد ضاوة</i> نيس اللجنة العلمية لقسم الاعلام الالي
Conseil Scientifique de la Faculté (ou de l'institut)	
Avis et visa du Conseil Scientifique : <i>Favorable</i>	
Date : <i>25/02/2024</i>	Présidente du Conseil Scientifique de la Faculté des Sciences <i>Pr. A B E D Hafida</i>
Doyen de la faculté (ou Directeur d'institut)	
Avis et visa du Doyen ou du Directeur :	
Date : <i>25/02/2024</i>	<i>عميد كلية العلوم</i> <i>د. لوارابي أحمد</i>
Conseil Scientifique de l'Université (ou du Centre Universitaire)	
Avis et visa du Conseil Scientifique :	
Date :	<i>استاذة بوزينة محمد</i> <i>جامعة البليدة 01</i>

VIII - Visa de la Conférence Régionale

(Uniquement à renseigner dans la version finale de l'offre de formation)