

Options AS

OPTION DATA SCIENCE		
Data science	volume horaire	Crédits
SEMESTRE 5		
UE : Système d'information	70	5
Bases de données avancées	20	1,5
Datawarehouse	30	2
SIG Avancé	20	1,5
UE : Statistique décisionnelle 1	85	7
Natural Language Processing (NPL)	25	2
Big data et Cloud computing	20	2
Webscraping	20	1,5
Marketing	20	1,5
SEMESTRE 6		
UE : Statistique décisionnelle 2	75	6
Machine Learning Optimization	30	2,5
Programmation web	25	2
Visualisation de données avancée	20	1,5
Total Data science	230	18

Base de données avancées

Semestre : 5

Volume horaire : 20h

UE : Système d'Information

Crédit : 2

1. Description et but du cours

1.1. Description du cours :

Le cours de bases de données avancées de manière générale, présente de nouvelles connaissances,et de nouvelles techniques offertes par les SGBD dits actifs. Il permet aussi de donner aux apprenants la capacité de savoir régler des problèmes de stockage et de manipulation de données complexes.

1.2. But du cours

Ce cours se veut essentiellement pratique et vise à donner aux futurs statisticiens :

- La possibilité d'approfondir leurs connaissances du modèle relationnel,
- La capacité d'implémentation des SGBD actifs afin d'être prêt à affronter les réalités des nombreuses problématiques de gestion de données sur le terrain.

2. Compétences professionnelles et niveau de maîtrise attendu

2.1. Objectifs du cours

L'objectif général du cours est d'étendre le concept de base de données préalablement étudié, à certains autres aspects et concepts auxquels devront certainement faire face les étudiants en fin de formation une fois sur le marché de l'emploi. On se rassurera que l'étudiant puisse suivre efficacement la schématisation qui part de la conception d'une base de données relationnelles au développement des déclencheurs en passant par la maîtrise des requêtes à travers les langages de manipulation et de définition de données DML et DDL.

2.2. Lien entre le cours et les autres cours du programme

Le cours de base de données avancées requiert :

- Des connaissances sur le modèle relationnel et le langage SQL au minimum ;
- Des notions de programmation (paradigme impératif) ;
- La possibilité d'avoir accès à un ordinateur avec un système d'exploitation Windows.

2.3. Eléments de compétences propres au cours et niveau de maîtrise attendu

Les éléments de compétences qui doivent être développés au cours de la session ainsi que leur niveau de maîtrise attendu sont précisés dans le tableau ci-dessous.

Éléments de compétence	Objectifs spécifiques (niveau de maîtrise attendu)
Installation et aperçu des outils de travail.	Savoir installer les différents applicatifs autour desquels sera bâti l'enseignement.
Présenter <i>Windesign</i> et <i>Mysql</i>	Connaitre et savoir utiliser <i>Windesign</i> comme outil de modélisation ; Savoir utiliser et configurer le SGBD <i>Mysql</i> comme outil de travail principal.
Présenter le concept de réPLICATION de bases de données	Brève introduction et implémentation de la réPLICATION de bases de données monodirectionnelle Master/slave

3. Situations d'enseignement-apprentissage (SEA)

Le cours sera structuré en trois parties :

- La première constituée à 9% du volume horaire sera accordée à l'installation et la présentation des applicatifs (outils) de travail.
- La deuxième constituée à 82% du volume horaire sera accordée à l'enseignement pratique et l'apprentissage des rouages liés à l'utilisation du SGBD *Mysql* ainsi que la mise en évidence de tous les concepts nouveaux.
- La dernière partie constituée à 9% du volume horaire, représente une amorce au concept de sécurité des bases de données.

4. Évaluation des apprentissages

4.1. Modalités d'évaluation sommative

Au regard de l'aspect essentiellement pratique de cet enseignement, une seule évaluation sous forme de mini-projet est préconisée.

Modalités	Individuel	En équipe	Total (%)
Evaluation écrite			
Exposés			
Evaluation pratique		de 2 ou 3 étudiants	100
Total (%)			100

Contenu du cours :

Chapitre 1 : Installation des outils de travail (2h)

1. Installation de *Xampp* ;
2. Installation de *Windesign* ;
3. Installation de *Notepad++*.

Chapitre 2 : Quelques rappels SQL de base (2h)

1. Rappel sur les concepts de base de la *modélisation des données* ;
 - *Rappel sur les concepts liés aux : Entités – Attributs – Propriétés* ;
 - *Rappel sur les concepts d'Association et de CIF* ;
 - *Rappel sur la mise en œuvre d'un MCD*,
2. Rappel sur l'utilisation de la clause SQL *SELECT* ;
3. Rappel sur l'utilisation des *jointures* SQL ;

Chapitre 3 : Autres types de requêtes avec MySQL (2h)

1. Les requêtes encadrées par le langage DML ;
 - *Utilisation des clauses de manipulation de données offertes par MySQL* ;
2. Les requêtes encadrées par le langage DDL ;
 - *Utilisation des clauses de définition de données offertes par MySQL*.

Chapitre 4 : Les moteurs de stockage (2h)

1. Présentation de MyISAM ;
2. Présentation de InnoDB ;

3. Présentation de MEMORY ;
4. Présentation de CSV.

Chapitre 5 : Les transactions (5h)

1. Définition du concept de transaction ;
2. Conception d'un modèle de données transactionnel : Ex transaction bancaire ;
3. Phase pratique : (Modélisation du scénario, création des différentes tables).

Chapitre 6 : Les Triggers (5h)

1. Généralités ;
2. Qu'est-ce qu'un trigger ?
3. Que fait un trigger ?
4. Syntaxe générale de création d'un trigger ;
5. Phase pratique : Création et manipulation des triggers.

Chapitre 7 : La réPLICATION de base de données en temps réel avec MySQL (2h)

1. Généralités et contexte ;
2. Les prérequis ;
3. La configuration du **MASTER** ;
4. La configuration du **SLAVE** ;
5. Les tests de synchronisation.

Contrôle des connaissances : 2 contrôles écrits

Bibliographie :

- 1 – Gardarin G., « Bases de données objet et relationnelle », Edition Eyrolles, 1999 (ISBN : 2-212-09060-9) ;
- 2 – Miranda S., Ruols A., « Client-serveur : concepts, moteur SQL et architecture parallèles », Eyrolles, 1994.

Webographie :

- 1 - <http://dev.mysql.com>
- 2 - <http://erci.no-ip.com/origine/static/php/yogui-php-cours.pdf>

Datawarehouse

Semestre : 5

Volume horaire : 30h

UE : Système d'Information

Crédit : 2

Objectifs du cours : Ce cours permet aux étudiants de se familiariser au concept et aux domaines d'application de des datawarehouse

Contenu du cours :

Introduction

Chapitre 1 : Généralités

- Historique du datawarehouse
- Notion de datawarehouse
- Caractéristiques des données d'un datawarehouse
- Architecture d'un datawarehouse

Chapitre 2 : OLAP et Analyse multidimensionnelle OLAP

- OLAP
- 12 règles de l'OLAP
- Opérations OLAP

Chapitre 3 : Modélisation et conception d'un datawarehouse

- Étude préalable
- Modélisation
- Alimentation
- Extraction

Chapitre 4 : Pratique sur Talend et mysql

Contrôle des connaissances : 1 contrôle écrit + 1 projet

Bibliographie :

SIG Avancé

Semestre : 5

Volume horaire : 20h

UE : Système d'Information

Crédit : 1.5

Objectifs du cours : Après un premier cours d'initiation au SIG, ce cours a pour but d'approfondir les connaissances dans le domaine des SIG en abordant le traitement et l'analyse des données satellitaires.

Contenu du cours :

Chapitre 0 : Introduction et Quelques notions de base

Chapitre 1 : Manipulation de données raster

Chapitre 2 : Traitement sur QGIS

Chapitre 3 : Calcul de statistiques et analyse multicritère

Chapitre 4 : Collecte et l'exportation de données

Chapitre 5 : Traitement des données

- Modification d'une table
- TP sur la manipulation, extraction des données
- Imputations de données
- Sélection par la localisation

Contrôle des connaissances : contrôle et projet à rendre

Bibliographie :

Natural Language Processing (NPL)

Semestre : 5

Volume horaire : 25h

UE : Statistique décisionnelle 1

Crédit : 2

Objectif : Le traitement du langage naturel est un domaine de l'intelligence artificielle concerné par le traitement systématique des langages humains. C'est maintenant une technologie courante utilisée dans une grande variété de produits comme l'assistant vocal, les moteurs de recherche, les systèmes de recommandation. Ce cours est une introduction exhaustive à la PNL. Nous couvrirons l'ensemble du pipeline de traitement de la PNL, du prétraitement et de l'apprentissage de la représentation à l'apprentissage supervisé spécifique à une tâche.

Contenu du cours :

Chapitre 1 : Les bases de la PNL

- Qu'est-ce que la PNL ?
- Spécificité de la PNL
- Comment aborder un problème de PNL ?

Chapitre 2 : Techniques d'apprentissage de la représentation pour le traitement automatique du langage naturel.

Chapitre 3 : Méthodes d'apprentissage en profondeur pour la PNL

- les techniques de Deep Learning utilisées en PNL.
- les principes de conception et de formation.
- le Multi-Layer-Perceptron (MLP), les architectures récurrentes (RNN et LSTM) et l'architecture des transformateurs.

Chapitre 4 : Modélisation du langage

- Introduction à la modélisation du langage.

Chapitre 5 : Étiquetage et classification des séquences

Étiquetage et classification des séquences avec des modèles d'apprentissage en profondeur tels que les modèles et les transformateurs basés sur RNN.

Chapitre 6 : Génération de séquences

- Génération de séquences à l'aide d'encodeurs-décodeurs.

Bibliographie :

- [Benoit Sagot 2022], Algorithms for speech and natural language processing, MVA course Material
- [Warren Weaver, 1949] Memorandum on Translation
- [Weizenbaum, 1966] Eliza
- [Dryer, Matthew S. & Haspelmath, Martin (eds.) The WALS]
- Distributional Semantics, Evert 2015
- Foundations of Distributional Semantics, Evert & Lenci 2009
- Stanford Information-Retrieval Book
- Miller et. al 1985: WordNet
- Mikolov et. al 2013: Word2vec
- Grave & Joulin, ENSAE 2018-2019

Big data et cloud computing

Semestre : 5

Volume horaire : 20h

UE : Statistique décisionnelle 1

Crédit : 2

Objectifs du cours :

Ce cours a pour principaux objectifs : introduire les notions et les domaines d'application des big data et leur lien avec la statistique classique

Contenu du cours :

Introduction :

Chapitre 1 : Généralités

- Définition et phase d'évolution du big data
- Characteristics of big data
- Types of big data and Examples of big data
- Importance des big data
 - Benefits of big data
 - Rules of big data

Chapitre 2 : Cloud computing

- Qu'est-ce que le cloud computing
- Concepts de base
- Services du cloud computing
- Déploiements du cloud computing
- Avantages / risques du cloud computing
- Caractéristiques du cloud computing
- Phase pratique

Logiciels : Google cloud Platform (GCP), Amazon Web Services (AWS)

Contrôle des connaissances : Projet à rendre

Bibliographie :

Webscraping

Semestre : 5
Volume horaire : 20h

UE : Statistique décisionnelle 1
Crédit : 1,5

Objectifs du cours :

Ce cours a pour principaux objectifs : permettre ...

Contenu du cours :

Chapitre 1: Eléments de rappel

- Rappel sur les techniques de collectes de données
- Rappels sur le web

Chapitre 2 : Définitions

- définition du web Scraping
- contexte d'utilisation
- cadre juridique

Chapitre3 : Prérequis techniques

- rappels sur la programmation
- rappels sur python pour le web scraping
- comprendre le html pour le web scraping
- comprendre le CSS pour le web scraping
- comprendre le JavaScript pour le web scraping

Chapitre 4 : Fonctionnement technique du Web Scraping

Chapitre 5 : Web scraping avec BeautifulSoup

Chapitre 6 : Web scraping avec : Scrapy et Sélénum

Chapitre7 : Web scraping avec les API

Chapitre8 : Quel outil de web scraping choisir ?

Chapitre9 : Cas pratique

Projet web scraping, exemple avec l'indice des prix

Webographie :

- <https://www.youtube.com/watch?v=JwSS70SZdyM>
- <https://www.youtube.com/watch?v=l26J4oJmsD0>
- <https://www.youtube.com/watch?v=3WqIq4q0XCc&list=PLrdkQ44yPrk91oiVirN6cYRnCTHF1GUIv&index=14>
- <https://www.ionos.fr/digitalguide/sites-internet/developpement-web/web-scraping-avec-python/>
- <https://www.youtube.com/watch?v=kwfswajC0vI&list=PLrdkQ44yPrk91oiVirN6cYRnCTHF1GUIv&index=4>

Marketing

Semestre : 5
Volume horaire : 20h

UE : Statistique décisionnelle 1
Crédit : 1,5

Objectifs du cours :

- Sensibiliser les étudiants à l'esprit « marketing » et à ses enjeux contemporains.
- Familiariser les étudiants avec la démarche marketing.
- Fournir les concepts, les outils d'analyse et les modèles spécifiques qui sont essentiels à la prise de décisions stratégiques afin d'élaborer une stratégie marketing : analyse & diagnostic, réflexion et choix stratégiques.
- Choisir et formuler la bonne stratégie marketing. Permettre une appropriation des concepts au travers d'études de cas

Chapitre 1 : Généralités sur le marketing

- Introduction
- Le cadre et la définition du marketing
- L'évolution du marketing
- Les notions clefs du marketing
- Les techniques, outils et les champs d'application
- Les limites du marketing

Chapitre 2 : Diagnostique stratégique marketing

Chapitre 3 : Diagnostic stratégique marketing : interne & externe, SWOT

Chapitre 4 : Segmentation-ciblage- Positionnement

Chapitre 5 : Choix d'une stratégie marketing

Contrôle des connaissances :

Bibliographie :

Machine learning Optimization

Semestre : 6

Volume horaire : 30 h

UE : Statistique décisionnelle 2

Crédit : 2,5

1. Description et but du cours

1.1. Description du cours

L'apprentissage-machine est le domaine d'étude qui donne aux ordinateurs la possibilité d'apprendre sans être explicitement programmé" [SAMUEL, 2000]. Ce cours explique comment utiliser un algorithme d'apprentissage-machine dont la performance et le comportement dépendent des données.

1.2. But du cours

Ce cours vise à donner aux étudiants les techniques et outils nécessaires du machine Learning, un sous-domaine de l'IA.

2. Compétences professionnelles et niveau de maîtrise attendu

2.1. Objectifs du programme

L'objectif général de ce cours se décline en plusieurs objectifs spécifiques comme suit :

1. Présenter les différents paradigmes du Machine Learning
2. Présenter les réseaux de neurones artificiels
3. Présenter les réseaux neuronaux convolutifs
4. Présenter l'apprentissage par transfert

2.2. Liens entre le cours et les autres cours du programme

Pour suivre convenablement ce cours, l'étudiant devra au préalable avoir suivi plusieurs cours dont :

1. Algèbre
2. Analyse
3. Statistique descriptive
4. ProbabilitéAlgorithmique
5. Programmation (R, Python)

Les connaissances acquises dans ce cours seront fortement utilisées dans des cours portant sur le Deep Learning (exclusivement), la vision par ordinateur et le NLP.

2.3. Éléments de compétences propres au cours et niveau de maîtrise attendu

Les éléments de compétences qui doivent être développés au cours de la session ainsi que leur niveau de maîtrise attendu sont précisés dans le tableau ci-dessous :

Eléments de compétence	Objectifs spécifiques (niveau de maîtrise attendu)
Présenter les différents paradigmes du Machine Learning	Savoir mettre en œuvre les techniques d'apprentissage supervisé et non supervisé
Présenter les réseaux de neurones artificiels	Savoir configurer et utiliser un réseau de neurones
Présenter les réseaux neuronaux convolutifs	Savoir configurer et utiliser un réseau de neuronal convolutif
Présenter l'apprentissage par transfert	Savoir configurer et utiliser un réseau de neurones pré-entraîné

3. Méthodes d'enseignement

Le cours consiste en des présentations complétées par des sessions pratiques et des travaux de groupe.

4. Évaluation des apprentissages

4.1. Modalités d'évaluation sommative

Les évaluations devront mettre l'accent sur le volet pratique. A cet effet, l'enseignant n'hésitera pas à donner des travaux de maison pour permettre aux étudiants de s'entraîner. L'évaluation finale sera un projet à faire individuellement ou en groupe selon la convenance du professeur.

5. Plan du cours

Chapitre 1 : Les paradigmes du Machine Learning (4h)

1. Place du Machine Learning dans l'IA
2. Apprentissage non supervisé
3. Apprentissage supervisé
4. Apprentissage par renforcement
5. Problématiques du Machine Learning

Chapitre 2 : Les réseaux de neurones artificiels (4h)

1. Architectures des réseaux de neurones
2. Entrainement d'un réseau de neurones
3. Techniques de régularisation

Chapitre 3 : Les réseaux neuronaux convolutifs (4h)

1. Les opérations mathématiques de *convolution* et de *pooling*
2. Architectures d'un CNN
3. Optimisation des performances d'un CNN

Chapitre 4 : L'apprentissage par transfert (4h)

1. Modèles pré-entraînés
2. Utilisation de la technique d'apprentissage par transfert

Chapitre 5 : Étude de cas : Analyse d'une problématique spécifique de Machine Learning (4h)

Contrôle de connaissances : 1 projet et/ou 1 contrôle écrit

Bibliographie :

1. Hefin Ioan Rhys, "Machine Learning with R, the tidyverse, and mlr"
2. François Chollet, "Deep Learning with Python", 2nd Edition, Manning

Programmation Web

Semestre : 6
Volume horaire : 25h

UE : Statistique décisionnelle 2
Crédit : 2

Objectifs du cours :

Ce cours a pour principaux objectifs : permettre aux étudiants de maîtriser les bases de la programmation sous HTML et sous Python

Contenu du cours :

Chapitre 1 : HTML

- Définition
- Structure d'une page HTML
- Les balises
- Création de page web
- Création de tableau
- Création de formulaires

Chapitre2 : Notion client/serveur : architecture et exemple

- Serveur web
- Javascript
- Les feuilles de style (CSS)
- Serveur web – utilisation d'IIS

Chapitre3 : programmation web Sous Python

- Installation des logiciels
- Utilisation de python avec pycharm
- Création de base de données SQL
- Interconnection de l'application avec SQLite

Contrôle des connaissances : Projet à rendre

Bibliographie :

Visualisation de Donnée Avancée

Semestre : 5

Volume horaire : 20h

UE : Statistique décisionnelle 2

Crédit : 1,5

Objectifs du cours : Se familiariser avec le logiciel Tableau

Introduction

- Introduction à l'outil logiciel : tableau public
- Chargement de données
- Manipulation des données

Chapitres 1 :

- Génération de visualisations
- Exploitation des différents indices visuels

Chapitres 2 :

- Exploitation des filtres
- Exploitation des animations

Chapitre 3 :

- Construction de tableaux de bords
- Étude du choix des paramètres pour une intégration dans un tableau de bord
- Exportation des projets

Contrôle des connaissances : Projet à rendre

Bibliographie :