

LICENCE PROFESSIONNELLE

**METIERS DE L'INFORMATIQUE
SYSTEMES D'INFORMATION
ET GESTION DES DONNEES
(SIGD)**

PROGRAMME DE LA FORMATION – PARCOURS BIG DATA

TABLE DES MATIERES

UE0 : Enseignements fondamentaux – SIGD-DATA	3
<i>EF1 : Algorithmique – Programmation Python (24H)</i>	3
<i>EF2 : Bases de données (12h)</i>	3
<i>DATA-EF3 : Statistiques (12H).....</i>	3
UE1 : Modules « environnement » de la spécialité SIGD-DATA	4
<i>SIGD-ENV1: Anglais (24H).....</i>	4
<i>SIGD-ENV2 : Expression – Communication d’entreprise (24H)</i>	4
<i>SIGD-ENV3 : Jeu d’entreprise (18H)</i>	4
<i>SIGD-ENV-4 : Gestion de projet informatique (30h)</i>	5
<i>SIGD-ENV5 : Droit des données et de la sécurité (12H)</i>	5
<i>DATA-ENV6: Big-Data et architectures associées Intro (18h).....</i>	5
UE2 : Modules « informatique-Developpement » - SIGD-DATA	6
<i>SIGD-MI2 : Développement Front-End (24H).....</i>	6
<i>SIGD-MI3 : Développement WEB EN PHP (24h).....</i>	6
<i>DATA-MI4 : Utilisation des frameworks pour les traitements distribués (24H).....</i>	6
<i>DATA-MI5 : Programmation pour l’interprétation des données (24H).....</i>	7
<i>DATA-MI6 : Sécurité des données et contrôle d’accès aux Big-Data (24h).....</i>	7
UE3 : Modules « informatique » Spécialité Statistiques SIGD-DATA	8
<i>DATA-INFO1 : Modélisation statistique (24H)</i>	8
<i>DATA-INFO2 : Statistiques exploratoires et multidimensionnelles (30h)</i>	8
<i>DATA-INFO3: Data mining (30H)</i>	9
<i>DATA-INFO4 : Data mining avancé (24H).....</i>	9
UE4 : Modules « Gestion des données » - SIGD DATA.....	10
<i>SIGD-GD1 : principes des BD noSQL (30H)</i>	10
<i>SIGD-GD2 : Informatique décisionnelle, datawarehouse (30h)</i>	10
<i>SIGD-GD3 : Visualisation des données (24h)</i>	11
<i>SIGD-GD4 : Extraction de données (18H).....</i>	11

EF1 : ALGORITHMIQUE – PROGRAMMATION PYTHON (24H)

Pré-requis : Quelques notions de programmation

Objectifs - Compétences minimales :

- Rigueur et l'imagination nécessaires pour une bonne maîtrise de la programmation en général,
- Construction d'algorithmes classique sur les vecteurs et sur les listes chaînées,
- Connaissance d'un certain nombre d'algorithmes classiques utilisables dans de nombreuses applications,
- Bases de la programmation en Java avec les structures de données et de contrôles classiques ainsi que gestion des exceptions pour mettre en œuvre les algorithmes étudiés
- Bases de la programmation par objets (encapsulation, classe, héritage...),
- Maîtrise de l'utilisation d'un IDE pour la programmation Java.

Contenu : Composants élémentaires des algorithmes, vecteurs et listes, programmation orientée objet, langage Java

Indications de mise en œuvre : Conception d'algorithmes avant implémentation et test en Java sous un IDE (Net-Beans ou Eclipse). Le contenu dispensé sera adapté au niveau de chaque étudiant.

EF2 : BASES DE DONNEES (12H)

Pré-requis : aucun

Objectifs – Compétences minimales : acquérir les notions théoriques et pratiques sur le modèle relationnel.

Contenu :

- Fondements des bases de données relationnelles,
- SQL,
- Programmation et administration de bases de données.

Indications de mise en œuvre : travaux pratiques avec utilisation du SGBD Postgres, création et interrogation de bases de données SQL.

DATA-EF3 : STATISTIQUES (12H)

Pré-requis : Notions de statistiques descriptives. Logiciel de traitement statistique des données (par exemple R).

Objectifs – Compétences minimales :

Faire découvrir ce qu'est une analyse de données à partir d'un jeu de données réelles : pré-traitement, exploration, analyse, visualisation, interprétation des résultats.

Contenu :

- statistiques descriptives : type de données, tableaux de résumés statistiques, représentations graphiques des données, indicateurs de liaisons.
- nettoyage et qualité de données

Indications de mise en œuvre : Mise en pratique sur R.

SIGD-ENV1: ANGLAIS (24H)

Pré-requis : Quelques connaissances et des compétences en anglais général et en anglais de spécialité en informatique.

Objectifs – Compétences minimales : Remise à niveau de certains points grammaticaux en cas de besoin, entraînement systématique à l'expression orale, langue de spécialité: anglais scientifique (informatique) et professionnel.

Contenu : Un programme d'enseignement mixte axé sur les 4 compétences langagières et sur la langue de spécialité (scientifique et professionnelle).

Indications de mise en œuvre : en salle banalisée (une pédagogie basée sur des tâches communicatives), travail en groupes et pédagogie par projets.

SIGD-ENV2 : EXPRESSION – COMMUNICATION D'ENTREPRISE (24H)

Pré-requis : aucun

Objectifs – Compétences minimales : familiariser les étudiants aux bases de la communication.

Contenu :

- Savoir se présenter efficacement en contexte professionnel (présentation minute personnalisée en fonction d'un objectif précis, gestion du stress)
- Les facteurs-clés de succès d'une réunion au top (en amont, pendant et après la réunion)
- Ecrire court pour être lu (mettre du relief dans un contenu, le structurer)
- La communication interpersonnelle (écoute active et reformulation, jeux de pouvoir, gestion des conflits pour débloquer une situation, savoir formuler une demande claire)
- L'animation d'équipe (identifier les profils de personnalités et les différentes postures au sein d'un groupe)

Indications de mise en œuvre : Conduite de réunion, Jeux de stratégie, Etude des mentalités, Jeux de rôles, Entretiens, Prise de décision, explications.

SIGD-ENV3 : JEU D'ENTREPRISE (18H)

Pré-requis : Notion de gestion

Objectifs – Compétences minimales : Permettre un apprentissage des techniques de gestion via un simulateur de gestion et concevoir des outils de gestion (Excel).

Contenu : Le module est plus particulièrement orienté vers les problèmes comptables et financiers, mais comme les principales fonctions de l'entreprise sont représentées (marketing, achat, production, personnel, finances), il permet de bien illustrer la prise en compte de leur interdépendance dans la mise en œuvre des décisions. Une importance particulière est attachée à la gestion prévisionnelle.

Indications de mise en œuvre : Les participants sont regroupés en équipes de 4 à 5 personnes, qui constituent les directions des différentes sociétés qui seront en concurrence dans le même environnement économique.

SIGD-ENV-4 : GESTION DE PROJET INFORMATIQUE (30H)

Objectifs – Compétences minimales : Grilles d'analyse et de lecture permettant de comprendre le domaine de la gestion de projet dans le domaine de l'informatique et de mieux appréhender les missions en entreprise et d'enrichir l'activité de restitution attendue dans le cadre du point entreprise et du rapport de fin d'année.

Contenu :

- Cadrage : Les éléments d'analyse de l'existant et de contexte à prendre en compte dans le cadre de la gestion des projets
- Analyse et gestion du risque : La notion de risque et de contrainte, criticité et priorisation, les différentes stratégies de mitigation. Le suivi des risques et l'évaluation de l'efficacité des plans de mitigation.
- Choix des modèles de développement : Les critères d'analyse des modèles de développement, présentation des principaux modèles. Lien vers l'ordonnancement et la planification, présentation des outils.
- Qualité : La qualité dans les projets logiciels : qualité ergonomique et qualité logicielle, analyse des critères SQuaRE, lien avec le test de logiciels.

Indication de la mise en œuvre : Chaque concept sera présenté puis assorti d'une fiche de guidance à remplir dans le cadre de leur alternance qui seront l'objet de séance de travail et remédiation puis les différents concepts seront approfondis et illustrés dans le cadre de séances encadrées et de séminaires professionnels.

SIGD-ENV5 : DROIT DES DONNEES ET DE LA SECURITE (12H)

Objectifs – Compétences minimales :

Afin de tenter de cerner les problématiques juridiques liées aux technologies informatiques, il est nécessaire de garder à l'esprit que nous sommes en présence non seulement d'un vecteur de communication entraînant des questionnements sur la qualification juridique de mécanismes comme le téléchargement de contenu ou le streaming, mais aussi en présence d'un objet nouveau protégé juridiquement et susceptible de faire l'objet d'atteinte.

Contenu : Dans le cadre de ce cours plusieurs axes thématiques seront abordés en partant de sujets d'actualité.

- La propriété intellectuelle appliquée au domaine de l'informatique et des réseaux.
- Le droit de la presse (orienté sur les règles applicables en matière de la liberté de la presse, de la responsabilité éditoriale et des délits de presse). Cette thématique suppose en amont une sensibilisation au domaine des libertés individuelles et publiques (approche historique et constitutionnelle).
- Le droit des données orienté "données personnelles" (loi informatique et Libertés GDPR, loi pour une République numérique), thématique supposant une sensibilisation aux notions et aux problématiques relatives à la protection de la vie et de l'intimité.
- La cybercriminalité et cyberdélinquance : application des règles spécifiques (fraude informatique) et adaptation du domaine du droit pénal général.

DATA-ENV6: BIG-DATA ET ARCHITECTURES ASSOCIÉES INTRO (18H)

Pré-requis : Aucun

Objectifs – Compétences minimales : Ce module vise à introduire le domaine du Big Data.

Contenu :

Il en définit les caractéristiques (volumétrie, vitesse, variété ...), et le caractérise par rapport aux domaines connexes tels que l'informatique décisionnelle et le datamining. Il situe les problématiques spécifiques du Big Data tout en décrivant les principales solutions technologiques existantes. Il en définit les caractéristiques (volumétrie, vitesse, variété ...), et le caractérise par rapport aux domaines connexes tels que l'informatique décisionnelle et le datamining. Il situe les problématiques spécifiques du Big Data tout en décrivant les principales solutions technologiques existantes (BD noSQL, Framework de traitement distribués, extraction et analyse de données textuels et multimédia...). Il présente aussi les domaines applicatifs fréquents du Big Data (Marketing, Presse, Gestion de production, Web ...).

Indications de mise en œuvre : Il s'appuie sur des cas concrets avec des présentations industrielles et une étude de cas réalisée dans l'environnement Spark avec une application du Machine Learning. Une seconde application est proposée dans l'environnement H2o avec une initiation au Deep Learning.

UE2 : MODULES « INFORMATIQUE-DEVELOPPEMENT » - SIGD-DATA

SIGD-MI2 : DEVELOPPEMENT FRONT-END (24H)

Pré-requis : EF1

Objectifs : Aborder les principales notions et technologies pour la création d'une page Web coté client: structurer un document, élaborer son rendu graphique, intégrer différents médias interactifs (techniques du web responsive)

Contenu :

- Principaux concepts de HTML 5, CSS3, le Framework Bootstrap 4
- Programmation dynamique avec Javascript, jQuery, AJAX. Interaction avec des données persistantes (JSON/XML).

Mise en œuvre : TP fil rouge à réaliser au fil des concepts abordés

SIGD-MI3 : DEVELOPPEMENT WEB EN PHP (24H)

Pré-requis : EF-1,

web, et structurer son code selon le pattern Modèle-Vue-Contrôleur (MVC)

Contenu : EF-2.

Objectifs - Compétences minimales : maîtriser la programmation orientée objet en PHP7 pour accéder à une base de données depuis un serveur

- Programmation Web et orientée objet en PHP7
- Accès à une base de données depuis un serveur web PHP7 (couche PDO)
- Structuration de code selon le pattern Modèle-Vue-Contrôleur (MVC)

Indications de mise en œuvre : Développement d'une petite application (base de données + site web)

DATA-MI4 : UTILISATION DES FRAMEWORKS POUR LES TRAITEMENTS DISTRIBUES (24H)

Pré-requis : Programmation Java

Objectifs – Compétences minimales : Acquisition des fondamentaux des infrastructures d'analyse de données massives de type Map/Reduce (Hadoop, Spark, Flink) :

- principes de stockage et traitement distribué de données massives sur cluster,
- structuration des données en tableaux clé/valeur
- développement de scripts d'analyse à base d'opérateurs parallèles (map, reduce, ...),
- organisation et analyse de différents types de données faiblement structurées (pages web, log chins, grands graphes de réseaux sociaux,...)

Contenu :

Le cours sera structuré en une partie cours qui présentera les principes des architectures map/reduce et une partie travaux pratiques où les étudiants seront amenés à structurer des jeux de données et développer des scripts d'analyses sur différents types de données.

Indications de mise en œuvre La programmation se fera en Scala et Java au-dessus d'une infrastructure Hadoop.

DATA-MI5 : PROGRAMMATION POUR L'INTERPRETATION DES DONNEES (24H)

Prérequis

- Notions d'algorithmique et programmation
- Notions de statistiques

Objectif : savoir programmer des algorithmes en lien avec les traitements statistiques dans le cadre d'utilisation de packages adéquates

Contenu – Indication de mise en oeuvre

Traitement des données sous python - anaconda

- Importation de données et liaison aux bases de données
- Utilisation de Pandas pour le traitement des données, dataframes
- Introduction à NumPy et SciPy, matrices et vecteurs
- Création automatisée de rapports de traitements

Données textuelles, temporelles et/ou géographiques

- Visualisation de données (numériques, textuelles, temporelles ou géographiques)
- Modèles probabilistes sous python (Bayes simplifié, graphes) - pgmpy

DATA-MI6 : SECURITE DES DONNEES ET CONTROLE D'ACCES AUX BIG-DATA (24H)

Prérequis : Vision et compréhension globale de l'organisation d'une société

Objectifs – Compétences minimales : Comprendre les principes de base et les méthodes de mise en oeuvre des briques élémentaires de sécurité des données en termes d'authentification, d'autorisation et de protection des données.

Contenu

- Principes de sécurité : Disponibilité, Intégrité & Non Répudiation, Confidentialité, Authentification & Autorisation
- Domaines d'application de la sécurité (Données, Infrastructures, etc)
- Menaces : menaces actuelles, application au Big Data
- Gestion des identités et des accès (Authentification & Autorisation)
- Protection des données (Intégrité, Confidentialité) : Classification des données, Signature et chiffrement, Outils et infrastructures, Anonymisation
- Protection des infrastructures et Applications (Disponibilité, Intégrité, Confidentialité)

Indications de mise en oeuvre : Une approche pratique sera mise en oeuvre pour illustrer et appliquer les différentes briques élémentaires au travers d'exemples et/ou projets

UE3 : MODULES « INFORMATIQUE » SPECIALITE STATISTIQUES SIGD-DATA

DATA-INFO1 : MODELISATION STATISTIQUE (24H)

Pré-requis

- Statistiques descriptives
- Notions élémentaires de probabilités et de statistique inférentielle

Objectifs – Compétences minimales

- Comprendre, mettre en œuvre et interpréter les modèles de régression linéaire simple et multiple, régression logistique
- Comprendre les limites de l'application des méthodes classiques aux données massives.

Contenu

- Modèle linéaire : régression simple, multiple, tests associés, tests de sous-modèles, sélection de variables (lasso, ...), bagging, analyse des résidus, ...
- Régression logistique : modélisation, algorithme récursif, lien avec le data-mining

Indications de mise en œuvre / Langages et/ou outils

- Logiciel R

DATA-INFO2 : STATISTIQUES EXPLORATOIRES ET MULTIDIMENSIONNELLES (30H)

Pré-requis : DATA-EF3, Statistiques descriptives, Notions élémentaires d'algèbre linéaire

Objectifs – Compétences minimales : Comprendre, mettre en œuvre et interpréter les résultats d'analyses exploratoires utilisées dans le traitement des données massives en tenant compte de leur hétérogénéité et de la possibilité d'avoir des données manquantes ; en particulier, savoir proposer des visualisations interprétables.

Contenu :

- Récupération et mise en forme d'un jeu de données réel. Initiation à R.
- Analyse en composantes principales (ACP) (avec et sans données manquantes)
- Classification Ascendante Hiérarchique (CAH)
- Suivant le temps : Analyse factorielle des correspondances (AFC)

Indications de mise en œuvre / Langages et/ou outils : logiciel R (packages FactomineR, missMDA, fastcluster, NMF...)

DATA-INFO3: DATA MINING (30H)

Pré-requis : Statistiques exploratoires et multidimensionnelles

Objectifs – Compétences minimales

- Comprendre, mettre en œuvre et interpréter les résultats des méthodes standards de classification utilisée pour le traitement des données massives
- Comprendre la construction et l'évaluation d'une règle de classification

Contenu

- Intro à la classification (supervisée, semi-supervisée et non supervisée)
- Méthodes de classification non supervisée : K-means, classification via EM
- Méthodes de classification supervisée : K plus proches voisins, arbres de décision (CART), forêts aléatoires, échantillons d'apprentissage, de validation et de test. Lien avec le cours de modélisation statistique.
- Sélection de modèle : critère AIC, BIC et ICL, validation croisée.
- Extraction de règles d'association

Indications de mise en œuvre / Langages et/ou outils

- Logiciel R
- Library Mllib
- Logiciel Orange

DATA-INFO4 : DATA MINING AVANCÉ (24H)

Pré-requis : DATA-INFO3, EF2

Objectifs – Compétences minimales : Connaître et comprendre les grands principes et techniques pour la fouille de texte et de séries temporelles.

Contenu

- Fouille de textes
 - définition des données textuelles
 - catégorisation documents (classification supervisée de textes) : réseaux de neurones, SVM et naïve bayes
 - partitionnement de documents (classification automatique) : K-means et CAH
- Fouille de séries temporelles
 - définition et indice descriptif d'une série temporelle
 - mesure de similarité
 - classification de séries temporelles supervisée et non supervisée

UE4 : MODULES « GESTION DES DONNEES » - SIGD DATA

SIGD-GD1 : PRINCIPES DES BD NOSQL (30H)

Pré-requis : Notions de SGBD classique

Objectifs - Compétences minimales :

- Faire découvrir l'univers du monde NoSQL
- Comprendre les différences majeures entre le SQL et le NoSQL
- Etre capable de proposer une technologie NoSQL adaptée à un problème donné

Contenu :

- Rappel de quelques notions de SQL (principe ACID)
- Historique du NoSQL
- Transactions et cohérence des données en NoSQL (Théorème CAP)
- Schémas de données dans les bases NOSQL avec les 4 grandes familles présentées (paires clés-valeurs, bases orientés documents, bases orientées colonnes, bases orientées graphiques)

Indications de mise en œuvre/langages et/ou outils : Etude de la base NoSQL MongoDB

SIGD-GD2 : INFORMATIQUE DECISIONNELLE, DATAWAREHOUSE (30H)

Pré-requis : module EF-2

Objectifs – Compétences minimales : concevoir, mettre en place et utiliser un système décisionnel.

Contenu :

- Bases de données réparties, fédérées, intégrées,
- Entrepôts de données et systèmes d'aide à la décision,
- Conception d'un entrepôt de données,
- Alimentation d'un entrepôt de données : rétro-conception de bases de données sources, ETL
- Composants logiciels pour la mise en place d'un entrepôt de données,

Indications de mise en œuvre : Retro-conception de bases opérationnelles, Conception, mise en place d'un entrepôt de données et alimentation d'un entrepôt, Exploitation des données d'un entrepôt : présentation dimensionnelle des données avec Qlik et PowerBI ;

SIGD-GD3 : VISUALISATION DES DONNEES (24H)

Pré-requis : notions élémentaires de géométrie et d'algèbre linéaire, maîtrise d'un langage de script (python/perl/etc.) connaissance des technologies web.

Objectifs – Compétences minimales :

- Présenter les fondements de la visualisation d'information, ainsi qu'un panorama des techniques de l'état de l'art applicables à différents types de jeux de données
- Mener à bien un projet de datavisualisation à partir d'un jeu de données

Contenu : Le cours est une introduction au domaine de la visualisation des données :

- Principes généraux de visualisation des données
- Les différents moyens de visualisation des données
- Les outils disponibles et les approches de visualisation
- La visualisation interactive

Indications de mise en œuvre Langages et/ou outils : Projet de data-visualisation à partir d'un jeu de données et réalisation d'un tutoriel autour d'une technique de datavisualisation.

SIGD-GD4 : EXTRACTION DE DONNEES (18H)

Pré-requis : Notions de programmation.

Objectifs – Compétences minimales :

- Savoir extraire des données issues de sources hétérogènes.
- Savoir transformer des données
- Savoir utiliser un outil ETL

Contenu : Utilité et principes de l'extraction, la transformation et du chargement de données.

Indications de mise en œuvre : travaux pratiques (extraction, transformation, chargement) avec l'ETL Talend.