

DATA SCIENCE FOR INDUSTRY

Modules Data Science

Machine Learning et Applications (Master MFD)- MSAMA	Lun	6 ECTS	17.5 ECTS
Introduction à la vision artificielle (cours ENS Ulm) - ENSVI	Mar	3 ECTS	
Architecture Big Data (GI) - ABDAT	Mer	1.5 ECTS	
Apprentissage Automatique pour la Recherche Opérationnelle (GI) - ROAPA	Mer	2 ECTS	
Deep Learning (IMI) - DEEPL	Mer	2 ECTS	
IA pour la Collaboration Homme-Machine (GI) - IACHM	Ven	3 ECTS	

Modules Complémentaires

Analyse de Cycle de Vie et écoconception (GMM) - COECO	Mer	1.5 ECTS	9.5 ECTS
Science du Changement Climatique (SEGF) - SGFCH	Lun	1,5 ECTS	
Ingénierie Système (GI) - INSYS	Jeu	2 ECTS	
Management Stratégique des Entreprises (GI) - MSTR	Ven	3 ECTS	
Semaine d'ouverture : Intelligence Economique (GMM)	Sept	1,5 ECTS	

Dept. Génie Industriel
Dept. Ingénierie Mathématique & Informatique

Machine Learning & Applications (Master MFD) - MSAMA

Enseignants: Romuald Elie, Jérémie Jakubowicz et Jean-Yves Audibert

Emploi du temps: lundi de 8h30 à 11h30 à l'Ecole des Ponts

Le but du cours est de présenter les principales méthodes théoriques de l'apprentissage statistique ainsi qu'un large spectre de leurs applications, en particulier pour la gestion de bases de données de taille importante. Le cours sera ponctué d'interventions de professionnels du monde de la donnée, qui viendront présenter des applications opérationnelles de ces méthodes aux domaines de l'actuariat, la finance et le marketing web.

- Fondements théoriques de l'apprentissage statistique: notion de risque et de risque empirique
- Régression logistique et classification
- Dimension de Vapnik, choix de la base de régression
- Méthode des k plus proches voisins, convexification du risque
- Réseaux de neurones
- Mise à jour de pondération d'estimateurs et application en gestion de portefeuille.
- Exemples d'applications en Actuariat, Finance et marketing web.

Introduction à la Vision Artificielle (cours ENS Ulm) - ENSVI

Enseignants: Mathieu Aubry, Gül Varol, Karteek Alahari, Ivan Laptev

Emploi du temps: mardi de 9h00 à 12h00 à l'ENS Ulm

Ce cours présente les principes et les fondations techniques de la vision artificielle, un domaine scientifique dont le but est de doter les ordinateurs de la capacité d'interpréter le contenu des images numériques (photographies et vidéos). Le cours comprend des exercices de programmation.

- Formation des images : Modèles des appareils de prise de vue, de la lumière et de la couleur.
- Traitement d'image local : Filtres, détection de contours, caractéristiques visuelles, texture.
- Groupes de pixels : Méthodes de "clustering", régression, et segmentation.
- Plusieurs images : Géométrie multi images, stéréo, analyse du mouvement.
- Analyse de scène : Détection et reconnaissance de visages, sacs de caractéristiques visuelles pour la reconnaissance de catégories d'objets.

Architecture Big Data (GI) - ABDAT

Enseignant: Eloïse Gomez

Emploi du temps: mercredi de 8h30 à 11h15 à l'Ecole des Ponts (deuxième partie du semestre)

L'importance de l'utilisation des données n'étant plus à prouver, les entreprises se trouvent désormais face à un défi de taille : traiter plus de données, plus rapidement et à moindre coût. Comment sont gérés actuellement les projets d'accès aux données et comment faire pour améliorer cette gestion au quotidien ?

Le cours « Architecture Big Data » permet d'acquérir des connaissances de bases pour comprendre l'environnement dans sa complexité. L'objectif est de comprendre les concepts sous-jacents à la réalisation d'un environnement de Big Data et les différents outils disponibles et nécessaires dans cet univers. Les parties théoriques permettront de comprendre l'architecture des différents composants. Les travaux pratiques permettront de se familiariser avec cet environnement.

Les différentes technologies qui seront utilisées : Hadoop, spark, hive, Kubernetes, ans, gcp, introduction à la plateforme Dataiku.

A travers des cours théoriques et des travaux pratiques les élèves seront capables de :

- Comprendre les différents composants d'un projet big data et comment ils interagissent ensemble
- Comprendre les architectures distribuées
- Appréhender le paradigme big data et le mécanisme de gouvernance de données
- Utiliser les principales technologies associées : Hadoop, Spark, hive, docker, kubernetes
- Mettre un projet en production et réussir à monitorer les ressources utilisées
- Apprendre à utiliser l'environnement GCP (créer une instance, créer une bucket pour stocker de la donnée, lancer des jobs sur cloud run)
- Maîtriser les principaux concepts de Dataiku
- Utiliser DSS pour automatiser le calcul d'un modèle de prédiction

Apprentissage Automatique pour la Recherche Opérationnelle (GI) - ROAPA

Enseignant: Axel Parmentier

Emploi du temps: mercredi de 13h45 à 16h30 à l'Ecole des Ponts (première partie du semestre)

Le développement des systèmes d'information a démultiplié la quantité de données accessibles sur les processus industriels. Les départements de recherche opérationnelle de l'industrie exploitent ces données pour optimiser les processus industriels, avec notamment des objectifs de fiabilité et de maîtrise des risques. On pense par exemple à l'exploitation de données capteurs en maintenance prédictive, à l'anticipation de la demande (données web, etc.) et d'entonnoirs dans la production (données usine) dans la gestion de la supply chain, aux données météorologiques pour gestion des énergies renouvelables, ou encore aux prévisions de retard pour les compagnies de transport. Ces données étant abondantes, parfois parcellaires, et souvent intrinsèquement aléatoires, l'apprentissage automatique est la discipline adéquate pour le contrôle de l'incertitude dans leurs traitements. Cependant, la combinatoire importante des problèmes rencontrés fait que les industriels se limitent en général aux approches du type "recherche opérationnelle déterministe", sans être capables d'utiliser les outils de l'apprentissage automatique.

L'objectif de cours est d'introduire l'optimisation data driven et l'apprentissage structuré, deux théories mathématiques à la frontière de la recherche opérationnelle et de l'apprentissage automatique qui permettent d'aborder ces problématiques. On insistera sur la modélisation et la résolution de problématiques industrielles avec ces outils. Ce cours sera l'occasion d'introduire des outils mathématiques mobilisés par ces théories : processus gaussiens (cours de proba de 2A), RKHS et apprentissage automatique, modèles graphiques probabilistes et optimisation Bayésienne, apprentissage structuré.

A l'issue de ce cours, les étudiant devront :

- Connaître les concepts de base de l'apprentissage structuré, de l'optimisation Bayésienne, de l'optimisation data driven et des modèles graphiques probabilistes.
- Être capables de reconnaître les problèmes industriels pouvant être abordés par les outils précédents.
- Être capables de mettre en œuvre ces outils dans des situations simples.

Ce module contribue au développement des compétences suivantes : comprendre les réalités physiques, techniques et humaines : analyser, modéliser, diagnostiquer ; mobiliser les outils de la recherche et du développement et de l'innovation ; appréhender et gérer la complexité des systèmes.

Contenus et éléments de programme

- Apprentissage structuré pour la recherche opérationnelle
- Optimisation Bayésienne pour la recherche opérationnelle
- Applications à l'optimisation résiliente des processus industriels (ordonnancement, routing, network design, etc.)

Deep Learning (IMI) - DEEPL

Enseignant: Mathieu Aubry

Emploi du temps: mercredi de 16h45 à 19h30 à l'École des Ponts (première partie du semestre)

Les outils d'apprentissage profond ont pris ces dernières années une importance considérable dans les applications industrielles, au point que de très nombreux élèves de l'École des Ponts s'y trouvent confrontés pendant leur stage/année de césure ou même durant leurs projets de département. L'enjeu de ce module est donc de les amener à utiliser ces outils de manière efficace, critique et pertinente.

À l'issue du module, les élèves auront compris les composantes essentielles de l'apprentissage profond: (i) architectures ; (ii) fonctions de coûts ; (iii) optimisation.

Les élèves seront également sensibilisés à la pratique de l'apprentissage profond, et en particulier aux problèmes du sur/sous apprentissage, aux stratégies d'optimisation, aux sources d'erreur courantes aux pratiques courantes, aux outils d'analyse des résultats et à l'utilisation d'une librairie d'apprentissage profond moderne (PyTorch)

Le module sera organisé en trois blocs comprenant chacun un cours et un TP/TD :

- Introduction et bases des réseaux de neurones (perceptron, perceptron multi-couches, fonction de coût, backpropagation..). Le TP/TD visera à implémenter complètement des réseaux de neurones simples, sans utiliser de librairies particulière.
- Réseaux de neurones modernes et profonds (Adam, batch normalization, architectures résiduelles, data augmentation, fine-tuning, implémentation efficace...) Le TP/TD portera sur l'entraîner un modèle convolutionnel profond avec PyTorch, en mettant en relief les choix clés à effectuer.
- Thématiques avancées, éléments de référence sur un certain nombre de thèmes populaires (réseaux génératifs adversariaux/GANs, auto-encodeur variationnel/VAE, apprentissage non et faiblement supervisé, "one-shot", attention. Le dernier TP/TD visera à mieux comprendre intuitivement les réseaux.

IA pour la Collaboration Homme-Machine (GI) - IACHM

Enseignants: Sotiris Manitsaris, Alina Glushkova

Emploi du temps: vendredi de 13h00 à 15h45 à l'Ecole des Ponts

L'industrie 4.0 et l'usine intelligence se base sur de nouvelles approches technologiques permettant de faire converger les systèmes physiques et numériques pour l'amélioration de l'ensemble de la chaîne de valeur industrielle, du développement à la production en passant par la chaîne logistique. Dans ce contexte, des interactions homme-machine (IHM) sont déployées dans tous les secteurs d'activité où l'utilisation de machines automatiques nécessite une collaboration avec des individus. Ces IHM utilisent des capteurs, des robots, des systèmes sans fil et des logiciels pour collecter et analyser des données. Les informations issues de ces données permettent alors d'optimiser le management des opérations. Cette forte demande d'IHM impose de faire évoluer les systèmes industriels grâce à des technologies qui complètent et assistent le travail humain (robotique) et réduisent les accidents industriels causés par la défaillance des processus. Le développement de l'intelligence artificielle dans l'industrie permet d'améliorer ces interactions, notamment en entraînant les machines à anticiper et détecter des anomalies et à collaborer de manière plus efficace avec l'humain.

Dans un futur proche, les machines seront toutes dotées de diverses couches de perception qui seront alimentées par l'IA (par exemple véhicule autonome, drone, IoT, domotique, ordinateur etc.). Sur la base d'algorithmes d'apprentissage automatique ou profond, ces machines analyseront les intentions et les comportements humains (mouvements postures, expressions faciales, mesures physiologiques, etc.) et apprendront en continu. Ainsi, la machine se transforme en un collecteur de données capable d'alimenter des algorithmes sophistiqués pour analyser, reconnaître et anticiper des actions et comportements humains.

L'objectif du cours est de permettre aux étudiants de découvrir les étapes à suivre afin de concevoir et développer des systèmes d'IA permettant la collaboration Homme-Machine. A l'issue de ce module les étudiants seront capables de :

- Identifier les capteurs nécessaires à utiliser pour la collecte de données provenant de l'homme
- Traiter le signal enregistré et analyser des séries temporelles
- Identifier et adapter la méthode de Machine/Deep Learning en fonction du contexte et développer des couches de perception pour le système interactif humain-machine.
- Evaluer l'efficacité et la performance du système et de ses algorithmes en termes de modélisation, détection et reconnaissance de formes.

Contenus et éléments de programme

- Capture de données : enregistrement de données physiologiques à l'aide de « wearables » ou capture du mouvement avec des caméras RGB-D ;
- Analyse du signal : traitement de données en fonction de leur nature. Par exemple, segmentation de la scène détection de poses humaines sur une image RGB-D, prétraitement du signal issu de gyroscopes ou accéléromètres etc. ;
- Extraction des caractéristiques : exportation des descripteurs de mouvement et sélection des plus appropriés pour décrire le comportement humain (geste dynamique, postures statiques etc.) en fonction du contexte ;
- Représentation et modélisation : utilisation de modèles stochastiques pour la représentation du comportement humain ;
- Reconnaissance de formes : apprentissage statistique (Hidden Markov Models, State-Space, etc.) et architectures d'apprentissage profond (Convolutional Neural Networks, Long Short Term Memory, End-to-End, etc.) pour la reconnaissance précoce et l'alignement des séries temporelles ;

Analyse du cycle de vie et écoconception (GMM) - COECO

Enseignants: Laurent Brochard, Adélaïde Feraille

Emploi du temps: mercredi de 8h30 à 11h15 à l'Ecole des Ponts (première partie du semestre)

Face aux tensions croissantes sur la disponibilité des ressources naturelles et énergétiques, face aux enjeux environnementaux en général, face à une demande sociétale croissante et des réglementations en évolution, l'Industrie a aujourd'hui besoin d'ingénieurs de haut niveau sachant intégrer les enjeux environnementaux dans la conception des produits et systèmes, capables d'émettre un avis critique argumenté sur des positions et propositions dans le domaine. Ce module a pour objectif d'intégrer la dimension environnementale dans la démarche de conception de l'ingénieur.

Séance 1 : Introduction à l'éco-conception

L'éco-conception est une démarche de conception qui vise à remplir une fonction tout en limitant les conséquences environnementales. L'idée que les ressources de la planète sont limitées et que nous risquons de les exploiter à un rythme qui n'est pas soutenable n'est pas récente. L'économiste Thomas Malthus proposait déjà de restreindre la natalité à l'aube du 19e siècle alors que la population mondiale était dix fois moindre qu'aujourd'hui. Le progrès technique a beaucoup changé la donne, et, malgré 7 milliards d'habitants en 2010, les conditions de vie ne se sont pas dégradées pour autant. Néanmoins, la pression environnementale exercée par l'homme sur la planète commence à être mesurée scientifiquement, à l'image du phénomène de changement climatique.

Séance 2 : Indicateurs environnementaux

L'évaluation environnementale repose sur des indicateurs dont l'objectif est de quantifier l'impact sur l'environnement d'un objet, d'une activité ou bien encore d'une population. Les indicateurs environnementaux sont utiles en ACV pour l'évaluation des produits et services, mais pas seulement. Les décideurs publics et les autorités sanitaires se servent des indicateurs pour élaborer des réglementations ou mettre en place des fiscalités 'vertes'. Les entreprises se servent des indicateurs afin d'identifier et combler leurs faiblesses en termes d'impact écologique et promouvoir leur image.

Séance 3 : Méthodologie ACV

L'Analyse de Cycle de Vie (ACV), qui fait l'objet de ce chapitre, est une méthodologie normalisée d'évaluation environnementale. Cette méthodologie s'applique aussi bien à un objet qu'à un service et constitue le paradigme de toute évaluation environnementale. Nous allons voir que l'ACV s'apparente à une 'comptabilité' environnementale.

Séance 4 et 5 : Cas d'étude dieseltgate et pratique de l'ACV avec openLCA

En septembre 2015, le scandale du dieseltgate a soulevé la question de la pollution due aux moteurs diesel. Au delà des questions judiciaires liées à la triche lors de l'homologation des véhicules se pose la question de la pertinence de la motorisation diesel vis-a-vis de la motorisation essence (et autres alternatives). Afin d'y voir plus clair, le but de ce cas d'étude est d'étudier les fondements de physico-chimie à l'origine des émissions des moteurs thermiques. Une seconde partie est dédiée à l'évaluation des impacts environnementaux associés avec un logiciel d'ACV open source (openLCA).

Séance 6 : Conception multicritère

La prise en compte de l'évaluation environnementale dans la conception ajoute une contrainte supplémentaire au concepteur dont l'approche traditionnelle se limite à un cahier des charges technique sous la seule contrainte financière. Nous présentons différentes méthodes formelles pour aborder la conception multi-critère et discutons en particulier l'approche suggérée dans la norme ACV.

- Examen

Science du Changement Climatique (SEGF) - SGFCH

Enseignants: Aglae Jezequel

Emploi du temps: lundi de 13h45 à 16h30 à l'Ecole des Ponts (première partie du semestre)

Le module " science du changement climatique " s'adresse aux étudiants de troisième année et est obligatoire pour les étudiants du parcours " Sustainable and Green Finance ". Son objectif est de donner à ces étudiants des bases en sciences du climat, afin qu'ils soient en mesure de comprendre les informations que ces sciences peuvent donner aujourd'hui, celles que l'on peut espérer obtenir dans un futur proche, et leurs limites. Les étudiants devront être en mesure à la fin de ce module de comprendre les informations essentielles présentées dans le groupe 1 du Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC), traitant des bases physiques du système climatique et du changement climatique.

Les quatre premières séances se présenteront sous la forme de cours magistraux et seront l'occasion de faire le tour des bases de climatologie qui alimentent notamment les rapports des groupes 1 et 2 du GIEC, telles que détaillées ci-dessous :

Séance 1: Qu'est-ce que le climat ?

Séance 2: Variabilité climatique : du cycle journalier aux cycles glaciaires.

Séance 3: La modélisation du climat

Séance 4 : L'influence de l'homme sur le climat

Les deux dernières séances seront consacrées à un travail collectif des étudiants permettant d'évaluer leur appropriation des concepts présentés lors des premières séances, mais également d'approfondir trois sujets plus précis. Les étudiants seront répartis en trois groupes, responsables d'organiser l'un des débats listés ci-dessous lors de la dernière séance.

Débat 1 : Le concept de " point de bascule " est-il utile pour décrire et communiquer les futurs climatiques possibles ?

Débat 2 : Les événements extrêmes singuliers doivent-ils être attribués à l'influence humaine ?

Débat 3 : Est-il pertinent de mettre l'accent sur le consensus scientifique pour encourager les politiques climatiques ?

Ingénierie Système (GI) - INSYS

Enseignants: Jean Luc Wippler

Emploi du temps: jeudi de 8h30 à 11h15 à l'Ecole des Ponts (première partie du semestre)

La conception, l'amélioration et le maintien en condition opérationnelle de systèmes industriels complexes nécessite la maîtrise des concepts et des principes de l'ingénierie système. Ce cours a donc pour objectif de présenter les différentes étapes du processus d'ingénierie, de l'analyse du besoin jusqu'au retrait de service. A partir de cas pratiques ou d'un projet fil rouge, l'enjeu est d'exposer la démarche et les principes fondamentaux de la conception d'une architecture de systèmes complexes.

A l'issue de ce module les étudiants seront capables de :

- Connaître les enjeux et principes fondamentaux de l'ingénierie système.
- Mettre en application les différentes étapes du processus d'ingénierie système depuis l'analyse du besoin jusqu'au retrait de service.
- Construire une vision système et maîtriser l'ingénierie des exigences.
- Concevoir et valider une architecture système (intégration, vérification, validation, qualification).
- Modéliser des systèmes complexes, notamment avec SysML.

Ce module contribue au développement des compétences suivantes : comprendre les réalités physiques, techniques et humaines : analyser, modéliser, diagnostiquer ; mobiliser les outils de la recherche et du développement et de l'innovation ; appréhender et gérer la complexité des systèmes.

Contenus et éléments de programme

- Construire une vision système et maîtriser l'ingénierie des exigences : Qu'est-ce qu'un système ? Contribution des produits et services au système.
- Ingénierie système : définition et principales étapes du processus.
- Cycle de vie : de la genèse de l'idée au retrait de service et au démantèlement du produit.
- Analyse fonctionnelle : exigences fonctionnelles et non fonctionnelles, contraintes.
- Fondamentaux de l'ingénierie système : conception de l'architecture, vérification et validation système, maîtrise des évolutions du système.
- Appréhender les fondamentaux de l'architecture d'un système.
- Introduction à l'ingénierie dirigée par les modèles (Model Based System Engineering).

Management Stratégique des Entreprises (GI) - MSTRA

Enseignants: Benjamin Lehiany

Emploi du temps: vendredi de 8h30 à 11h15 à l'Ecole des Ponts

Dans un contexte de mondialisation croissante et d'intensification de la pression concurrentielle, les entreprises doivent de plus en plus prendre des décisions qui ne relèvent plus seulement d'un management opérationnel à court ou moyen terme. Face à un environnement en constante mutation, elles doivent prendre des décisions à long-terme afin de conserver une longueur d'avance et de favoriser une création de valeur pérenne.

Ce cours a donc pour objectif de former les élèves ingénieurs aux principes théoriques, aux méthodes et aux outils du management stratégique.

A l'issue de ce module les étudiants seront capables de :

- Conduire un diagnostic stratégique.
- Etre capable d'évaluer des stratégies.
- Maitriser la conception de choix stratégique.
- Etre capable de concevoir un déploiement stratégique.
- Formuler la stratégie d'une organisation de manière synthétique.
- Distinguer les caractéristiques de la stratégie au niveau de la direction générale, au niveau de chaque domaine d'activité ou au niveau opérationnel.

Ce module contribue au développement des compétences suivantes : comprendre les réalités physiques, techniques et humaines : analyser, modéliser, diagnostiquer ; mobiliser les outils de la recherche et du développement et de l'innovation ; appréhender et gérer la complexité des systèmes.

Contenus et éléments de programme

- Diagnostique stratégique (analyse de l'environnement, capacité stratégique, intention stratégique).
- Choix stratégiques (stratégies par domaine d'activité, diversification, gestion de portefeuille d'activités, stratégies internationales, innovation, entrepreneuriat, alliance et partenariats).
- Déploiement stratégique (évaluation des stratégies, processus stratégiques, stratégie et organisation),