**Guide pédagogique**

***Module « Sciences des données »***

***Option IASD – 8.5*** *(4 crédits ECTS)*

***Place du module et enjeux***

Ce module propose un enseignement approfondie sur les sciences des données : quels traitements basés sur des modèles statistiques et la théorie des probabilités permettent de "faire parler" ces données. A COMPLETER.

**Teaching guide and syllabus**

***« Data Sciences » module***

***DSAI option – 8.5*** *(4 ECTS credits)*

***Subject matter importance and associated issues***

A TRADUIRE

**Responsable : Gérard Dray**

Téléphone : 04 34 24 62 XX

Courriel : gerard.dray@mines-ales.fr

|  |  |  |  |
| --- | --- | --- | --- |
| **ENSEIGNEMENTS ACADÉMIQUES** | **Volume**  **horaire** | Détail des  coefficients | Crédits |
| **Sciences des données**   * Introduction à l'analyse de données * Statistiques et probabilités avancées * Validation, visualisation, restitution | **50 h**  5  35  10 | 1  4  2 | 4 |

**Matière 1 :**

|  |  |
| --- | --- |
| **Titre de la matière : Introduction à l'analyse de données** | |
| **Code : 2IA-iasd-8.5.1** | **Titre** **du module :** « Sciences des données » |
| **Semestre :** S8 | **Cursus de rattachement :**Département 2IA, option IASD |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Heures présentiel | Heures  total | Cours | TD | TP | Projet | Contrôles | Travail personnel | Coef /module | ECTS |
| 5 | 7 | 5 |  |  |  |  | 2 | 1 | 0.57 |

|  |  |
| --- | --- |
| **Titre** | Introduction à l'analyse de données |
| **Résumé** | L'option IASD du département forme des ingénieurs analystes de données. Cette introduction positionne l'analyse de données en regard de différentes problématiques. Une introduction au langage R est également proposé car c'est le langage de prédilection de ce domaine. |

|  |  |
| --- | --- |
| **Responsable** | Gérard Dray – LGI2P/IMT Mines Alès |
| **Équipe enseignante** | Gérard Dray – LGI2P/IMT Mines Alès  Nicolas Sutton-Charani – LGI2P/IMT Mines Alès |

|  |  |
| --- | --- |
| **Mots-clés** | Analyse de données, langage R |
| **Prérequis** | ???... |

|  |
| --- |
| **Contexte et objectif général :**  Les techniques actuellement mises en avant en Intelligence Artificielle reposent sur l'utilisation de données, nombreuses et hétérogènes, qui peuvent être croisées, analysées, afin de faire émerger des comportements récurrents, d'expliquer certains phénomènes, voire de prédire certains faits. Cette introduction permet de décrire les contextes propices à ce genre d'approches et donner un inventaire des techniques utilisées dans le domaine. |
| **Programme et contenu :**   * 3h d'introduction à l'analyse de données * 2h d'introduction au langage R |
| **Méthode et organisation pédagogique :**  Les enseignements sont prévus pour 46 élèves.  Le découpage est prévu comme suit :   * 5h de cours (illustration avec le langage R) |
| **Acquis d’apprentissage visés :**  *Savoir, connaitre, être capable de (appliquer, synthétiser….)*  XX. |
| **Évaluation :** *???* |
| **Retour sur l’évaluation fait à l’élève :**3 semaines après la dernière séance. |
| **Support pédagogique et références :**  1 Polycopié |

**Matière 2 :**

|  |  |
| --- | --- |
| **Titre de la matière : Statistiques et probabilités avancées** | |
| **Code : 2IA-iasd-8.5.2** | **Titre** **du module :** « Sciences des données » |
| **Semestre :** S8 | **Cursus de rattachement :**Département 2IA, option IASD |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Heures présentiel | Heures  total | Cours | TD | TP | Projet | Contrôles | Travail personnel | Coef /module | ECTS |
| 35 | 50 | 14 | 18 |  |  | 3 | 15 | 4 | 2.28 |

|  |  |
| --- | --- |
| **Titre** | Statistiques et probabilités avancées |
| **Résumé** | L'option IASD du département forme des ingénieurs analystes de données. Cet enseignement a pour objectif de renforcer les compétences théoriques avec un approfondissement des statistiques et de la théorie des probabilités. |

|  |  |
| --- | --- |
| **Responsable** | Nicolas Sutton Charani – LGI2P/IMT Mines Alès |
| **Équipe enseignante** | Nicolas Sutton Charani – LGI2P/IMT Mines Alès  Abdelhak Imoussaten – LGI2P/IMT Mines Alès  Stefan Janaqi – LGI2P/IMT Mines Alès |

|  |  |
| --- | --- |
| **Mots-clés** | Statistiques, probabilités, statistiques bayesiennes, estimation et vraisemblance, réduction de dimension, tests |
| **Prérequis** | Notions de base des statistiques et probabilités. |

|  |
| --- |
| **Contexte et objectif général :**  Les enseignements de l'option IASD (Intelligence Artificielle et Science des Donnes) et en particulier toutes les méthodes d'apprentissage automatique et de *business intelligence*, reposent sur des calculs statistiques et/ou des modèles probabilistes. Cet enseignement en fournit les fondements théoriques et donne des pistes pour leur mise en œuvre. |
| **Programme et contenu :**  Cet enseignement détaillera les chapitres suivants :   * Estimation et vraisemblance (3h de cours et 3h de TD) * Mélange de lois normales et algorithme EM – Espérance-maximisation (3h de cours et 3h de TD) * Statistiques Bayésiennes (3h de cours et 6h de TP) * Réduction de dimensionnalité (4h de cours et 4h de TD) * Tests (1h de cours 2h de TD) |
| **Méthode et organisation pédagogique :**  Les enseignements sont prévus pour 46 élèves en promotion complète, avec la possibilité d'un enseignant supplémentaire pour certaines séances de mise en pratique.  Le découpage est prévu comme suit :   * 14h de cours (illustration avec le langage R) * 18h de TD * 3h de contrôle sur table |
| **Acquis d’apprentissage visés :**  *Savoir, connaitre, être capable de (appliquer, synthétiser….)*  XX. |
| **Évaluation :***Types d’épreuves et répartition des coef : devoir sur table (1)*  Épreuve sur table |
| **Retour sur l’évaluation fait à l’élève :** *mise à disposition des corrections, consultation des copies etc :*  *Délais de correction des examens :…. (un maximum de 3 semaines est toléré pour un rendu de correction d’examens)*  3 semaines après la dernière séance. |
| **Support pédagogique et références :**  1 Polycopié |

**Matière 3 :**

|  |  |
| --- | --- |
| **Titre de la matière : Validation, visualisation, restitution** | |
| **Code : 2IA-iasd-8.5.3** | **Titre** **du module :** « Sciences des données » |
| **Semestre :** S8 | **Cursus de rattachement :**Département 2IA, option IASD |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Heures présentiel | Heures  total | Cours | TD | TP | Projet | Contrôles | Travail personnel | Coef /module | ECTS |
| 10 | 15 | 2 |  | 8 |  |  | 5 | 2 | 1.14 |

|  |  |
| --- | --- |
| **Titre** | Validation, visualisation, restitution |
| **résumé** | Un pan important de l'analyse de données concerne la restitution à l'utilisateur des différents résultats obtenus après traitement. Ce cours permet de présenter les différents modes d'affichage de ces résultats et d'outiller les étudiants avec la présentation de librairies logicielles dédiées. |

|  |  |
| --- | --- |
| **Responsable** | Gérard Dray – LGI2P/IMT Mines Alès |
| **Équipe enseignante** | Yan Holtz – Montpellier SupAgro  Doctorant en appui pour l'encadrement de TP |

|  |  |
| --- | --- |
| **Mots-clés** | Visualisation, Restitution… |
| **Prérequis** | Connaissance basique de R. |

|  |
| --- |
| **Contexte et objectif général :**  Les méthodes d’Intelligence Artificielle et de Science des Donnés sont toutes basées sur une compréhension de la donnée. Utiliser des méthodes de visualisations appropriées est une compétence indispensable pour mener ce travail à bien. |
| **Programme et contenu :**   * Introduction à la visualisation de données (2h de cours) * Introduction à ggplot2 (3h de TP) * Application à un jeu de données du choix des étudiants (2h de TP) * Construction et partage d'un rapport reproductible avec R Markdown et GitHub (3h de TP) |
| **Méthode et organisation pédagogique :**  Les enseignements sont prévus pour 46 élèves en promo complète. Les TP seront réalisés sur les ordinateurs personnels de ces derniers.  Le découpage est prévu comme suit :   * 2h de cours * 8h de mise en application en TP. * Évaluation : étude du dépôt de projet dans GitHub et du rapport en ligne. Les critères de notation seront en particulier : l'originalité, le choix des graphiques, la qualité du rapport et la clarté du code. |
| **Acquis d’apprentissage visés :**  Connaitre les bonnes pratiques de visualisation de données et les différents types de graphique. Savoir les construire à l’aide de R et ggplot2. Etre capable de les intégrer dans un rapport reproductible et les partager grâce à des outils de contrôle de version. |
| **Évaluation :** Types d’épreuves et répartition des coefficients : TP notés (2) |
| **Retour sur l’évaluation fait à l’élève :** 3 semaines après la dernière séance. |
| **Support pédagogique et références :** Supports de cours |

# Méthode et organisation pédagogique

Il s'agit d'un enseignement relativement classique avec une partie réalisée en cours magistral et une partie appliquée au travers de TP et Projets.

# Modalité d’évaluation

Le niveau d’acquisition des compétences sera évalué selon les exigences suivantes :

|  |  |
| --- | --- |
| N° indicateur | Indicateur |
| 1 | connaitre les savoirs formels et pratiques du socle des fondamentaux |
| 2 | Exploiter les savoirs théoriques et pratiques |
| 3 | Analyser, interpréter, modéliser, émettre des hypothèses, et résoudre |

*Répartitio*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Matière** | **Contrôle** | **Coefficients** | **Type de notation** | **Indicateurs évalués** | **Chapitres** |
| Introduction à l'analyse de données | **???** | 1 |  |  |  |
| Statistiques et probabilités avancées | Devoir sur table | 4 | Individuelle | 1 et 2 | Tous |
| Validation, visualisation, restitution | TP noté | 2 | En binôme | 3 | Tous |

# Engagement de l’étudiant, éthique et professionnalisme

La démarche éthique est définie dans le règlement intérieur de l’établissement. Chaque étudiant s’engage à en prendre connaissance et à la respecter.

*Obligation des cours* *:* Présence obligatoire pour tous à chaque séance

*Nombre d’heures estimées de travail personnel :* pour acquérir les compétences demandées, il est nécessaire que l’étudiant consacre minimum 45 min de travail personnel de compréhension et d’approfondissement par séance de cours.

*Nombre d’heures estimées de préparation aux travaux dirigés (TD) :* 1 à 2h.

*Pénalité pour retard*

Tout travail remis en retard sans motif valable peut être pénalisé de 1 point par jour de retard (notation effectuée sur 20).

# Équipe enseignante

|  |  |  |
| --- | --- | --- |
| **Nom** | **Domaine d’expertise** | **Courriel/Téléphone** |
| Gérard DRAY | Analyse et fouille de données | [gerard.dray@mines-ales.fr](mailto:gerard.dray@mines-ales.fr) / 04 34 24 62 xx |
| Abdelhak IMOUSSATEN | Analyse multicritère, Apprentissage des préférences, Théorie de l’incertain, Fusion de données, Opérateurs d’agrégation | [abelhack.imoussaten@mines-ales.fr](mailto:abelhack.imoussaten@mines-ales.fr)  04 34 24 62 64 |
| Stefan JANAQI | Data Science, Mathématiques appliquées | [stefan.janaqi@mines-ales.fr](mailto:stefan.janaqi@mines-ales.fr) 04 34 24 62 19 |
| Nicolas  SUTTON-CHARANI | Science de la Donnée, Apprentissage Automatique et Théories de l’Incertain | [nicolas.sutton-charani@mines-ales.fr](mailto:nicolas.sutton-charani@mines-ales.fr)  04 34 24 62 67 |
| Yann HOLTZ | Visualisation de données, Data science. | [Yan.holtz.data@gmail.com](mailto:Yan.holtz.data@gmail.com)  0678035324 |

|  |  |  |  |
| --- | --- | --- | --- |
| **ACADEMIC TEACHING** | **Teaching hours** | Coefficients | Credits |
| **Data Sciences**   * Introduction to data analysis * Advanced Statistics and Probability * Validation, Visualisation, Reporting | **50 h**  5  35  10 | 1  4  2 | 4 |

**Class 1**

|  |  |
| --- | --- |
| **Class title: Introduction to data analysis** | |
| **Code : 2IA-iasd-8.5.1** | **Module title:** « Data Sciences » |
| **Semester:** S8 | **Classification :** 2IA Department, DSAI option |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Hours of presence | Total hours | Lectures | Workshop | Labs | Project | Testing | Personal work | Coef /module | ECTS |
| 5 | 7 | 5 |  |  |  |  | 2 | 1 | 0.57 |

|  |  |
| --- | --- |
| **Title** | Introduction to Data Analysis |
| **Summary** |  |

|  |  |
| --- | --- |
| **Head** | Gérard Dray – LGI2P/IMT Mines Alès |
| **Teaching team** | Gérard Dray – LGI2P/IMT Mines Alès  Nicolas Sutton-Charani – LGI2P/IMT Mines Alès |

|  |  |
| --- | --- |
| **Key words** | Data analysis, R programming language |
| **Prerequisites** |  |

|  |
| --- |
| **Context and general objective:** |
| **Programme and contents:**  The course covers the following topics:   * introduction to data analysis (3h) * introduction to R programming language (2h) |
| **Method and pedagogic organisation:**  The course will be organized as follows for 46 students   * 5h of lectures * 1h written exam |
| **Targeted skills or knowledge :** |
| **Evaluation:** One hour written examination. |
| **Feedback made to the student** :  Exam copies within three weeks |
| **Teaching material and references :**  Lecture slides and reference textbooks |

**Class 2**

|  |  |
| --- | --- |
| **Class title: Advanced statistics and probability** | |
| **Code : 2IA-iasd-8.5.2** | **Module title:** « Data Sciences » |
| **Semester:** S8 | **Classification :** 2IA Department, DSAI option |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Hours of presence | Total hours | Lectures | Workshop | Labs | Project | Testing | Personal work | Coef /module | ECTS |
| 35 | 50 | 14 | 18 |  |  | 3 | 15 | 4 | 2.28 |

|  |  |
| --- | --- |
| **Title** | Advanced statistics and probability |
| **Summary** |  |

|  |  |
| --- | --- |
| **Head** | Nicolas Sutton Charani – LGI2P/IMT Mines Alès |
| **Teaching team** | Nicolas Sutton Charani – LGI2P/IMT Mines Alès  Abdelhak Imoussaten – LGI2P/IMT Mines Alès  Stefan Janaqi – LGI2P/IMT Mines Alès |

|  |  |
| --- | --- |
| **Key words** | Statistics, probability, Bayesian statisctics, … |
| **Prerequisites** |  |

|  |
| --- |
| **Context and general objective:** |
| **Programme and contents:**  The course covers the following topics:   * Maximum likelihood estimation (3h de cours et 3h de TD) * GMM (Gaussian Mixture Model) and Expectation–maximization algorithm (3h de cours et 3h de TD) * Bayesian Statistics (3h de cours et 6h de TP) * Reduction of dimensionality (4h de cours et 4h de TD) * Testing (3h de cours) |
| **Method and pedagogic organisation:**  The course will be organized as follows for 46 students   * 16h of lectures * 6h of labs * 3h written exam |
| **Targeted skills or knowledge:** |
| **Evaluation:** Three hour written examination. |
| **Feedback made to the student:** annotated exam copies within three weeks |
| **Teaching material and references:**  Lecture slides and reference textbooks |

**Class 3**

|  |  |
| --- | --- |
| **Class title: Validation, Visualisation, Reporting** | |
| **Code : 2IA-iasd-8.5.3** | **Module title:** « Data Sciences » |
| **Semester:** S8 | **Classification :** 2IA Department, DSAI option |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Hours of presence | Total hours | Lectures | Workshop | Labs | Project | Testing | Personal work | Coef /module | ECTS |
| 10 | 15 | 2 |  | 8 |  |  | 5 | 2 | 1.14 |

|  |  |
| --- | --- |
| **Title** | Validation, Visualisation, Reporting |
| **Summary** |  |

|  |  |
| --- | --- |
| **Head** | Gérard Dray – LGI2P/IMT Mines Alès |
| **Teaching team** | Yan Holtz – Montpellier SupAgro  Doctorant en appui pour l'encadrement de TP |

|  |  |
| --- | --- |
| **Key words** | Data visualization, Reproducibility, Reporting |
| **Prerequisites** | Basic knowledge on R |

|  |
| --- |
| **Context and general objective:**  Artificial intelligence and data science in general is always based on a previous step of data exploration. Visualizing the data the right way and creating a reproducible report is a crucial step in the pipeline. |
| **Programme and contents:**  The course covers the following topics:   * Introduction to data visualization (2h) * Introduction to ggplot2 (3h) * Application to a dataset of your choice (2h) * Building and sharing report with R markdown and Github (3h) |
| **Method and pedagogic organisation:**  The course will be organized as follows for 46 students   * 2h of lectures: * 8h of labs to learn the basis of ggplot2 and learn how to create and share a report * Evaluation |
| **Targeted skills or knowledge:**  Knowledge of different types of chart, Overview of dataviz best practices, Capacity to build basic charts with R and ggplot2, and to share them thanks to R Markdown and Github. |
| **Evaluation:** On project: analysis of the report shared online, including clarity of code. |
| **Feedback made to the student:** 3 weeks after end of Module. |
| **Teaching material and references:**  Slides will be provided. |

# Method and teaching organisation

Teachings will consist of lectures, seminars and labs.

# Testing procedures

The student’s level of knowledge acquisition will be evaluated according to the following points:

|  |  |
| --- | --- |
| N° Indicator | Indicator |
| 1 | To know the formal and practical knowledge constituting the foundation of a given field |
| 2 | Exploit theoretical and practical knowledge |
| 3 | Analyse, interpret, model, hypothesize and solve problems |

*Grading scheme:*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Class** | **Exam** | **Coefficients** | **Administration mode** | **Evaluated Indicators** | **Chapters** |
| Introduction to data analysis | Multiple choice exam | 1 | Individual | 1 | All |
| Advanced Statistics and Probability | Exam | 4 | Individual | 1, 2, 3 | All |
| Validation, Visualisation, Reporting | Project | 2 | ?? | 3 | All |

# Student commitments, ethics and professionalism

Expectations concerning ethics are defined in the establishment’s code of conduct. Each student is expected to know and respect the code of conduct.

*Obligatory presence in classes:* Students must attend all courses, seminars and labs.

*Estimated hours of personal study: in order to acquire the required learning level, the student is expected (must) to spend a minimum of 45min of personal study time per hour spent in class*.

*Estimated hours of preparation required for labs/Work Shop:*

1 to 2 hours

*Late penalties:* Late works are subject to penalties as follows: 3 points per day (ratings are between 0 and 20).

# Teaching team

|  |  |  |
| --- | --- | --- |
| **Name** | **Expertise** | **Email/Phone** |
| Gérard DRAY | Data analysis, Data maining | [gerard.dray@mines-ales.fr](mailto:gerard.dray@mines-ales.fr) / 04 34 24 62 xx |
| Abdelhak IMOUSSATEN | Multiple criteria decision analysis, Preference learning, Uncertainty theory, data merging, aggregation operators | [abelhack.imoussaten@mines-ales.fr](mailto:abelhack.imoussaten@mines-ales.fr)  04 34 24 62 64 |
| Stefan JANAQI | Data Science, applied mathematics | [stefan.janaqi@mines-ales.fr](mailto:stefan.janaqi@mines-ales.fr) 04 34 24 62 19 |
| Nicolas  SUTTON-CHARANI | Data Science, Machine Learning and Uncertainty theory | [nicolas.sutton-charani@mines-ales.fr](mailto:nicolas.sutton-charani@mines-ales.fr)  04 34 24 62 67 |
| Yan HOLTZ | Data visualization, data science | [Yan.holtz.data@gmail.com](mailto:Yan.holtz.data@gmail.com)  0678035324 |

# Approbation

Ce guide pédagogique entre en vigueur à compter du 07 janvier 2019

Il est porté à la connaissance des élèves par une publication sur le site de l'école

|  |  |  |
| --- | --- | --- |
| **Rédaction** | **Vérification** | **Validation** |
| L’enseignant responsable du module :  Gérard DRAY | Le responsable d’UE / de département :  Sylvie RANWEZ | Le directeur de l’école,  Pour le directeur et par délégation,  Le directeur de la DFA / de la DE : |