

## Introduction à la science des données

|                          |   |
|--------------------------|---|
| <b>Domaine</b>           | Ingénierie et Architecture                |
| <b>Filière</b>           | Informatique et systèmes de communication |
| <b>Orientation</b>       | Réseaux et systèmes (ISCR)                |
| <b>Mode de formation</b> | Plein temps                               |

### Informations générales

|                    |   |
|--------------------|---|
| Nom                | : Introduction à la science des données |
| Identifiant        | : ISD                                   |
| Années académiques | : 2020-2021, 2021-2022, 2022-2023       |
| Responsable        | : Andres Perez Uribe                    |
| Charge de travail  | : 120 heures d'études                   |
| Périodes encadrées | : 64 (= 48 heures)                      |

| Semestre    | E1 | S1 | S2 | E2 | S3 | S4 | E3 | S5 | S6 |
|-------------|----|----|----|----|----|----|----|----|----|
| Cours       |    |    | 32 |    |    |    |    |    |    |
| Laboratoire |    |    | 32 |    |    |    |    |    |    |

### Connaissances préalables recommandées

L'étudiant-e doit connaître et savoir utiliser les notions suivantes :

- bonnes notions de programmation procédurale
- formalisme mathématique, bases d'algèbre linéaire, calcul différentiel et matriciel, notions de géométrie analytique

### Objectifs

Ce cours sert d'introduction au domaine et aux bases méthodologiques de la science des données. A l'issue de cette unité d'enseignement, l'étudiant-e sera capable de:

- délimiter les problèmes qui relèvent de l'apprentissage automatique,
- distinguer les différents types de problèmes d'apprentissage et les paradigmes qui s'y rattachent (apprentissage supervisé et apprentissage non-supervisé)
- formaliser des énoncés de problèmes en termes de tâches d'apprentissage et mettre en place une méthode d'apprentissage appropriée,
- utiliser des bibliothèques du calcul scientifique conçues pour analyser les données disponibles et concevoir des modèles,
- savoir interpréter les résultats en faisant preuve d'esprit critique.

## Contenu et formes d'enseignement

Répartition des périodes indiquée à titre informatif.

### Cours: 32 périodes

- Introduction à la science des données: contexte, problématique, motivation, définition, domaines d'applications 2
- Données/ensembles de données: données, attributs ou variables, ensembles de données, types d'ensembles de données, données structurées et non structurées, classification des attributs, statistiques descriptives de base pour les attributs, représentations graphiques des statistiques descriptives élémentaires, principes de caractérisation des données 6
- Apprentissage supervisé: notions, méthodologie, types de problèmes, applications; exemples d'algorithmes de base (p.ex., régression linéaire, K-plus-proches voisins) 14
- Apprentissage non-supervisé: notions, méthodologie, types de problèmes, applications; exemples d'algorithmes de base 10

### Laboratoire: 32 périodes

- Utilisation des outils du calcul scientifique issus d'un langage de programmation de haut niveau (p.ex. Python et des bibliothèques telles que Numpy, Pandas, Matplotlib, Sklearn) 10
- Mise en place d'une approche d'apprentissage supervisé pour traiter des problèmes de classification et de régression: préparation des données, création des modèles (apprentissage), techniques de validation et sélection (validation croisée et hold-out), et évaluation de la qualité du modèle à l'aide de différents critères (F-score, matrice de confusion, courbes ROC) 16
- Mise en place d'une approche d'apprentissage non-supervisé et interprétation des résultats 6

## Bibliographie

- Azencott, Chloé-Agathe. Introduction au machine learning. Dunod, 2019.
- Domingos, Pedro. The master algorithm: How the quest for the ultimate learning machine will remake our world. Basic Books, 2015.
- VanderPlas, Jake. Python data science handbook: Essential tools for working with data. " O'Reilly Media, Inc.", 2016 (<https://jakevdp.github.io/PythonDataScienceHandbook/>)

## Contrôle de connaissances

**Cours** : l'acquisition de la matière de cet enseignement sera contrôlée au fur et à mesure par des travaux écrits individuels tout au long de son déroulement. Il y aura au moins 2 tests d'une durée totale d'au moins 2 périodes.

**Laboratoire** : ils seront évalués sur la base des rapports de manipulation, à 3 reprises au minimum.

**Examen** : L'atteinte de l'ensemble des objectifs de formation sera vérifiée lors d'un contrôle final commun écrit d'une durée de 90 minutes.

Matériel autorisé :

- Information communiquée directement par l'enseignant.

**Calcul de la note finale**

Note finale = moyenne cours x 0.3 + moyenne laboratoire x 0.2 + moyenne examen x 0.5