

ENG-209

**Data science pour ingénieurs avec Python**

Bouillet Eric Pierre, Pellet Jean-Philippe, Verscheure Olivier

| Cursus                           | Sem. | Type |
|----------------------------------|------|------|
| Génie électrique et électronique | BA3  | Obl. |
| HES - EL                         | H    | Opt. |

|                         |                     |
|-------------------------|---------------------|
| Langue d'enseignement   | français            |
| Crédits                 | 3                   |
| Session                 | Hiver               |
| Semestre                | Automne             |
| Examen                  | Pendant le semestre |
| Charge                  | 90h                 |
| Semaines                | 14                  |
| <b>Heures</b>           | <b>3 hebdo</b>      |
| Cours                   | 1 hebdo             |
| Projet                  | 2 hebdo             |
| <b>Nombre de places</b> |                     |

**Résumé**

Ce cours est divisé en deux parties. La première partie présente le langage Python et les différences notables entre Python et C++ (utilisé dans le cours précédent ICC). La seconde partie est une introduction aux outils, bibliothèques Python et méthodes collaboratives de data science.

**Contenu**

La première partie traite des spécificités du langage Python et passe en revue la syntaxe, les structures de données de base, les compréhensions de listes et de sets, la lecture et l'écriture de fichiers, l'usage de classes-données (dataclasses), et une introduction de base à un style de programmation plus fonctionnel.

La seconde partie expose les étudiants au processus complet de Data Science, depuis l'acquisition de données à l'extraction de connaissances appliquées à des problèmes réels. Cette partie se déroulera sur 8 semaines (à partir du 6e lundi de cours), selon les 5 modules suivants:

**1. Préparer son environnement interactif pour le Data Science**

- A# la dure: conda, git, Jupyter et docker
- Plus pratique: renku

**2. Manipulation et exploration des données avec la bibliothèque Python Pandas**

- Projet #1: Exploration de données CO2 provenant d'un réseau national de capteurs

**3. Visualisation de la donnée avec Matplotlib**

- Projet #2: Visualisation spatiotemporelle des données CO2 de ces capteurs

**4. Machine Learning (ML) avec Scikit-learn**

- Projet #3: Regroupement des capteurs de comportements similaires (clustering)
- Projet #4: Détection de comportements anormaux des capteurs (modélisation par régression)

**5. Communication des résultats à une audience non technique avec Voilà**

- Projet #5: Adaptation des projets 2, 3 et 4 en une application interactive web

**Mots-clés**

Programmation, Python, Data Science, Machine Learning

**Compétences requises**

**Concepts importants à maîtriser**

Notions de programmation de base:

- Structures de contrôles (instructions conditionnelles, boucles)
- Concepts de fonctions/méthodes
- Concepts de classes et d'instances

**Acquis de formation**

A la fin de ce cours l'étudiant doit être capable de:

- Modéliser en langage Python les éléments pertinents d'une situation problème
- Transposer les connaissances acquises en C++ vers Python
- Appliquer les constructions de haut niveau de Python pour créer des structures de données linéaires ou associatives
- Evaluer les avantages et inconvénient d'un langage haut niveau par rapport à un langage plus bas niveau
- Utiliser les bibliothèques Python standards pour le Data Science
- Concevoir des solutions applicables à des problèmes Data Science réels

**Compétences transversales**

- Utiliser une méthodologie de travail appropriée, organiser un/son travail.
- Accéder aux sources d'informations appropriées et les évaluer.
- Utiliser les outils informatiques courants ainsi que ceux spécifiques à leur discipline.
- Faire preuve d'esprit critique

**Méthode d'enseignement**

Séances ex cathedra avec exercices et travaux pratiques

**Travail attendu**

Participation aux cours, résolution d'exercices, réalisation du travail attendu en dehors de heures de cours

**Méthode d'évaluation**

Contrôle continu pendant le semestre (à définir selon conditions COVID-19) réparti selon:

- Examen mid-term (concluant la première partie): 1/3 de la note totale
- Examen final (sous forme de projet Data Science): 2/3 de la note totale

**Encadrement**

|                    |     |
|--------------------|-----|
| Office hours       | Non |
| Assistants         | Oui |
| Forum électronique | Oui |

**Ressources****Liens Moodle**

- <https://go.epfl.ch/ENG-209>