

Présentation

Code interne : GID9-DATAS

Description

Le module donne un premier socle en science des données et en intelligence artificielle appliquées aux géosciences et aux sciences de l'environnement. L'objectif n'est pas de former des data scientists en une semaine, mais de rendre chaque participant capable d'être un **interlocuteur efficace et critique** dans une chaîne de traitement de données : suivre un pipeline réel de la donnée brute à la sortie de modèle, identifier la méthode pertinente pour un problème donné, évaluer la qualité et les limites d'un modèle, et discuter à égalité avec des ingénieurs données, des modélisateurs et des analystes.

Pour les doctorants, qui travaillent le plus souvent sans support data dédié, la semaine est explicitement conçue comme le **point de départ d'un parcours autonome**, pas comme une formation complète : elle fournit le vocabulaire, la carte des méthodes et les réflexes de vérification qui permettent d'apprendre seul ensuite.

Semaine 1 (27 h, **ouverte aux participants extérieurs** — doctorants et professionnels ; enseignement et supports **en anglais** si nécessaire). Séances de cours de 1 h ou 2 h, **une séance unique par thématique** : mise en place de concepts et de réflexes de vigilance, l'approfondissement relevant du travail personnel ultérieur. S'y ajoutent trois retours d'expérience de praticiens en exercice (hydrologue en industrie ; data scientist, industrie de l'eau ; géoscientifique, industrie pétrolière et gazière) et une intervention d'un enseignant-chercheur invité sur les fonctions de transfert appliquées à la modélisation hydrogéologique. Évaluation : QCM.

Semaine 2 (étudiants ENSEGID uniquement). Travaux dirigés et projet appliqué sur des données de géosciences ou de sciences de l'environnement, mobilisant les méthodes vues en semaine 1. Évaluation : soutenance orale du projet.

Heures d'enseignement

CM	Cours Magistraux	27h
TD	Travaux Dirigés	21h

Pré-requis obligatoires

Exécuter et déboguer seul un script scientifique Python, manipuler des données tabulaires, anglais B2 (semaine 1 et supports en anglais si présence d'étudiants non francophones).

Syllabus

Semaine 1 – ENSEGID et externes :

- Introduction au monde de la donnée – 2 h
- Fonctions de transfert appliquées à la modélisation hydrogéologique (A. Pryet) – 2 h
- Analyse exploratoire des données – 1 h ; nettoyage et préparation des données : qualité, GIGO – 2 h
- Fondamentaux des bases de données – 1 h ; visualisation de données – 1 h
- Bases de l'apprentissage automatique – 2 h ; bases de l'apprentissage profond – 2 h
- Évaluation des modèles : surapprentissage, métriques, esprit critique face à l'IA – 2 h
- Jumeaux numériques – 2 h ; intégration de la physique dans les modèles d'IA – 2 h
- Retours d'expérience de praticiens : Extract Transform Load (ETL) et jeux de données temporels ; prédiction de ressources, hybridation, forêts aléatoires ; augmentation de données et reconnaissance automatique de motifs – 3 × 2 h
- Présentation des objectifs et des données des participants – 2 h

Semaine 2 – ENSEGID uniquement : TD et projet appliqué sur données de géosciences / environnement ; soutenance orale.

Modalités de contrôle des connaissances

Évaluation initiale / Session principale

Type d'évaluation	Nature de l'évaluation	Durée (en minutes)	Nombre d'épreuves	Coefficient de l'évaluation	Note éliminatoire de l'évaluation	Remarques
Contrôle Continu	Contrôle Continu					

Infos pratiques

Contacts

Responsable module

Anne-Laure Argentin

✉ Anne-Laure.Argentin@bordeaux-inp.fr