

PROG4 – Programming for Data Science (Programmation pour la science des données)

Course title - Intitulé du cours	PROG4 – Programming for Data Science (Programmation pour la science des données)
Level / Semester - Niveau / semestre	A1
School - Composante	Ecole d'Economie de Toulouse
Teacher - Enseignant responsable	Gaurav Dhar
Other teacher(s) - Autre(s) enseignant(s)	
Other teacher(s) - Autre(s) enseignant(s)	
Other teacher(s) - Autre(s) enseignant(s)	
Other teacher(s) - Autre(s) enseignant(s)	
Lecture Hours - Volume Horaire CM	15h
TA Hours - Volume horaire TD	
TP Hours - Volume horaire TP	15h
Course Language - Langue du cours	English
TA and/or TP Language - Langue des TD et/ou TP	English

Teaching staff contacts – Coordonnées de l'équipe pédagogique :

gaurav.dhar@tse-fr.eu

Course Objectives – Objectifs du cours :

This course introduces the concept of a pipeline in programming for data science. A pipeline is the implementation of data science processes. It organizes and executes the different stages of data preparation, analysis, modeling, and evaluation in order to transform raw data into insights, predictions, or decision-support tools.

The objectives of the course are as follows:

- Represent a machine learning pipeline as a sequence of organized and reproducible processes.
- Be able to implement an end-to-end pipeline, including data loading, model training, evaluation, and deployment.

Ce cours introduit le concept de **pipeline** en programmation pour la science des données. Un pipeline est la mise en œuvre des processus de science des données. Il organise et exécute les différentes étapes de préparation, d'analyse, de modélisation et d'évaluation afin de transformer des données brutes en connaissances, prédictions ou aides à la décision. Les objectifs du cours sont les suivants :

- Représenter un pipeline, c'est-à-dire une suite des processus organisées et reproductibles, en machine learning
- Être capable d'implémenter un pipeline depuis le chargement de données, ensuite l'entraînement de modèle, son évaluation et son déploiement

Prerequisites – Pré requis :

It is recommended to have some practice in a programming language and have an understanding of basic programming principles (variables, control structures, functions).

Il est recommandé d'avoir une expérience pratique avec un langage de programmation et de comprendre les principes fondamentaux de la programmation (variables, structures de contrôle, fonctions).

Practical information about the sessions – Modalités pratiques de gestion du cours :

Grading system – Modalités d'évaluation :

Exam and/or Project, Coding Assignments

Bibliography/references – Bibliographie/références :

VanderPlas, J. (2016). Python Data Science Handbook: Essential Tools for Working with Data. O'Reilly Media.

James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). An Introduction to Statistical Learning: With Applications in Python.

Deisenroth, M. P., Faisal, A. A., & Ong, C. S. (2020). Mathematics for Machine Learning. Cambridge University Press.

Session planning – Planification des séances

Distance learning – Enseignement à distance :