

5.00 credits

30.0 h + 15.0 h

Q2

Teacher(s)	Caelen Olivier ;Heuchenne Cédric ;
Language :	English
Place of the course	Louvain-la-Neuve
Prerequisites	Knowledge in programming (e.g. LDATS2030)
Main themes	This course covers the essential conceptual and practical components of applied data science. The major themes include: – Application domains of Data Science and real-world case studies illustrating end-to-end project development. – The standard life-cycle of a data science project: data acquisition, exploratory analysis, data cleaning and preprocessing, model development, evaluation, and deployment. – Fundamentals of supervised learning, with an emphasis on classification and regression problems. – Introduction to classical machine learning models, including decision trees, k-nearest neighbours, and neural networks... – Practical implementation using the Python scientific ecosystem and modern tools supporting reproducible pipelines and deployment.
Learning outcomes	At the end of this learning unit, the student is able to : With regard to the AA framework of the Master [120] in Data Science : Statistic, this activity contributes to the development and acquisition of the following AAs: • as first priority: 1.1 – 1.2 – 1.3 – 2.1 – 2.4 – 2.5 – 6.3 • as secondary: 4.1 – 4.2 – 4.4 – 5.3 – 5.6 – 6.1 ¹ With regard to the AA framework of the Master [120] in Statistics: General, this activity contributes to the development and acquisition of the following AAs: • as first priority: 1.3 – 2.2 – 2.5 – 3.3 – 5.3 • as secondary: 2.4 – 5.4 – 5.6 – 6.3
Evaluation methods	Group project and individual oral exam. The oral exam may include questions about the project. Both parts are mandatory to pass the course.
Teaching methods	The course combines Ex-cathedra course supported by slides with practical computer sessions in which students apply each concept using Python notebooks.
Content	1. Introduction to Data Science 1. Overview of the data science workflow: business understanding, data understanding, modelling, evaluation, deployment 2. Case-study 2. Data Extraction and Data Manipulation in Python 1. Introduction to the Python scientific ecosystem: NumPy, pandas, ... 2. Data loading from files and APIs (CSV, JSON, SQL) 3. Data preprocessing: missing values, feature engineering, encoding categorical variables, scaling and normalization ... 4. Exploratory data analysis and visualisation (matplotlib, seaborn,...) 3. Supervised Learning: Classification and Regression 1. Introduction to the Python machine learning ecosystem: sklearn & statsmodels 2. Classical machine learning models for tabular data: 1. k-nearest neighbours 2. decision trees and random forest 3. introduction to neural networks and PyTorch 4. Model Evaluation and Interpretability 1. Train/validation/test splits, cross-validation, performance metrics (accuracy, F1, ROC-AUC, MSE) 2. Feature importance and model explainability (SHAP, permutation importance)

	5. Introduction to Tools and Computing Environment 1. Jupyter, MLflow, Streamlit, FastAPI
Bibliography	"Hands#On Machine Learning with Scikit#Learn and PyTorch: Concepts, Tools, and Techniques to Build Intelligent Systems", Aurélien Geron, O'Reilly Media, ISBN -13: 979-8341607989 "Machine Learning with PyTorch and Scikit-Learn: Develop machine learning and deep learning models with Python", Sebastian Raschka, Yuxi (Hayden) Liu & Vahid Mirjalili, Packt Publishing, ISBN-13: 978-1801819312.
Faculty or entity in charge	LSBA

Programmes containing this learning unit (UE)				
Program title	Acronym	Credits	Prerequisite	Learning outcomes
Master [120] in Data Science : Statistic	DATS2M	5		
Master [120] in Statistics: Biostatistics	BSTA2M	5		
Master [120] in Linguistics	LING2M	5		
Master [120] in Environmental Bioengineering	BIRE2M	5		
Master [120] in Mathematics	MATH2M	5		
Master [120] in Actuarial Science	ACTU2M	5		
Master [120] in Statistics: General	STAT2M	5		
Master [120] in Chemistry and Bioindustries	BIRC2M	5		
Master [120] in Mathematical Engineering	MAP2M	5		
Minor in Statistics, Actuarial Sciences and Data Sciences	MINSTAT	5		
Certificat d'université : Statistique et science des données (15/30 crédits)	STAT2FC	5		