ESIROI
ÉCOLE D'INGÉNIEURS
UNIVERSITÉ DE LA RÉUNION

DEPARTMENT : *IT*
SEMESTER : *S8*

IDENTIFICATION

Course code : ESI-SPI-CI-IN4-S8-UE5-EC2
ECTS : 2

DURATION

Lectures :	4
Tutorial classes :	16
Labs :	0
Total :	20

Project :
Personnal work :

ASSESSMENT

Continuous assessments and evaluated practical work.

TEACHING MATERIALS

Slides, guided and practical work sheets.

LANGUAGE

English

CONTACT

Mathilde LEPETIT
mathilde.lepetit@univ-reunion.fr

Modifié le : 14 juillet 2026

UE : UE8-DATA

COURSE : Data Analysis : Methods and Applications

SCOPE OF THE COURSE

OBJECTIVES

Deepen and consolidate data science skills by exploring advanced analysis methods and their practical applications across various domains. Building on machine learning and deep learning foundations, this course emphasizes mastering the entire data project lifecycle — from data collection and preprocessing to result interpretation and production deployment — while integrating ethical, reproducibility, and model explainability considerations.

COMPETENCE AND SKILLS

- Design and manage a complete data analysis project by applying methods suited to the problem's nature, data volume, and quality
- Apply advanced methods for processing and analyzing tabular, temporal, textual, and multimedia data (images, signals) by selecting and justifying methodological choices
- Rigorously evaluate a data analysis model's performance, diagnose its limitations, and propose well-argued improvements
- Ensure reproducibility, traceability, and maintainability of a data analysis pipeline by applying MLOps best practices (versioning, testing, documentation)
- Interpret and explain complex model results using explainability methods (SHAP, LIME, Grad-CAM) and present them clearly to technical or non-technical audiences
- Integrate ethical and regulatory dimensions into data analysis system design : algorithmic biases, fairness, GDPR and RIA compliance

PROGRAM

- **Data project lifecycle and best practices :**
  - Data project scoping : defining objectives, identifying data, feasibility assessment.
  - Data lifecycle management : collection, storage, versioning, governance.
  - Reproducibility and traceability : MLflow, DVC, experiment tracking.
  - Development best practices : unit testing, documentation, CI/CD for data projects.
- **Advanced preprocessing and feature engineering :**
  - Handling missing, outlier, and imbalanced data : advanced strategies (SMOTE, multiple imputation).
  - Feature selection and construction : filter, wrapper, and embedded methods.
  - Encoding high-cardinality categorical variables ; adaptive normalization and standardization.
  - Data augmentation for images, text, and time series.
- **Temporal data and signal analysis :**

- Characterization and decomposition of time series : trend, seasonality, residuals.
- Classical models : ARIMA, SARIMA, Prophet.
- Deep learning approaches for time series : LSTM, Temporal Convolutional Networks (TCN), temporal Transformers.
- Anomaly detection in time series.
- Signal analysis : Fourier transform, spectrograms, wavelets.

— **Natural Language Processing (NLP) :**

- Text representation : bag-of-words, TF-IDF, word embeddings (Word2Vec, GloVe, FastText).
- Sequence models for NLP : RNN, LSTM, attention mechanisms.
- Transformers and pre-trained language models : BERT, CamemBERT, GPT — fine-tuning for specific tasks.
- Applications : text classification, sentiment analysis, information extraction, automatic summarization.

— **Multimedia data analysis (images and signals) :**

- CNN recap ; advanced architectures : EfficientNet, ViT (*Vision Transformer*).
- Object detection and segmentation : YOLO, Mask R-CNN.
- Transfer learning and *few-shot learning* for multimedia data.
- Audio data analysis : feature extraction (MFCC), sound classification, speech recognition.

— **Explainability and interpretability of models (XAI) :**

- Challenges of black-box models in deep learning ; *explainable AI* (XAI) concepts.
- Local methods : LIME, SHAP (*SHapley Additive exPlanations*).
- Global methods : feature importance, partial dependence plots (PDP).
- Visualizing CNN decisions : Grad-CAM, activation maps.

— **Ethics, bias, and regulatory compliance :**

- Sources of bias in data and models : selection bias, confirmation bias, representation bias.
- Algorithmic fairness metrics : disparate impact, equality of opportunity.
- Regulatory obligations : GDPR applied to data analysis systems, RIA classification, DPIA.
- *Responsible AI* principles : transparency, traceability, human oversight.

— **Production and deployment (MLOps) :**

- Architecture of an ML system in production : *feature store*, *model registry*, monitoring.
- Model deployment : REST API (FastAPI, Flask), containerization (Docker), orchestration (Kubernetes).
- Monitoring models in production : data drift, performance degradation.
- Continuous model updating and retraining.

ESIROI
ÉCOLE D'INGÉNIEURS
UNIVERSITÉ DE LA RÉUNION

BIBLIOGRAPHY

— Aurélien Géron - *Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow* - O'Reilly, 3rd ed., 2022
— Ian Goodfellow, Yoshua Bengio & Aaron Courville - *Deep Learning* - MIT Press, 2016 : `https://www.deeplearningbook.org`
— Chip Huyen - *Designing Machine Learning Systems* - O'Reilly, 2022
— Jacob Devlin et al. - *BERT : Pre-training of Deep Bidirectional Transformers for Language Understanding* - arXiv, 2018 : `https://arxiv.org/abs/1810.04805`
— Scott Lundberg & Su-In Lee - *A Unified Approach to Interpreting Model Predictions (SHAP)* : `https://arxiv.org/abs/1705.07874`
— Hugging Face Transformers documentation : `https://huggingface.co/docs/transformers`
— MLflow documentation : `https://mlflow.org/docs/latest/`
— FastAPI documentation : `https://fastapi.tiangolo.com`
— Papers With Code — State of the art by task : `https://paperswithcode.com`
— CNIL - *Recommendations on AI systems and GDPR* : `https://www.cnil.fr/en/artificial-intelligence`

REQUIREMENTS

— **DAT801 - Deep Learning (S8) :** CNN architectures, RNN, LSTM, optimization mechanisms, TensorFlow/PyTorch frameworks — direct and essential prerequisite for the entire course.
— **DAT701 - Machine Learning (S7) :** supervised and unsupervised algorithms, model evaluation, scikit-learn pipelines.
— **DAT701 - Massive Data Systems (S7) :** distributed architectures, Spark — required for large-scale processing pipelines.
— **Data Mining (S6) :** preprocessing, pattern extraction, introduction to clustering.
— **SHS704 - Law and Ethics in Computer Engineering (S7) :** GDPR, RIA, algorithmic biases — applied in the ethics and compliance section of the course.
— **Expected cross-disciplinary skills :**

  — Advanced Python programming : numpy, pandas, scikit-learn, PyTorch or TensorFlow
  — Proficiency in Git and code/data versioning tools
  — Basic knowledge of containerization (Docker) and application deployment
  — Ability to read and synthesize research papers in English
  — Scientific rigor : reproducibility, documentation, objective evaluation

ESIROI - Campus de Terre Sainte
40 Avenue de Soweto, 97455 Saint-Pierre, Cedex
Tel.: +262 262 81 83 01 / 02 ; www.esiroi-reunion.fr