ÉSIROI
ÉCOLE D'INGÉNIEURS
UNIVERSITÉ DE LA RÉUNION

DEPARTMENT : *IT*
SEMESTER : *S7*

IDENTIFICATION

Course code :	ESI-SPI-CI-IN4-S7-UE5-EC2
ECTS :	1

DURATION

Lectures :	4
Tutorial classes :	12
Labs :	8
Total :	24
Project :	
Personnal work :	

ASSESSMENT

Continuous assessments and evaluated practical work.

TEACHING MATERIALS

Slides, guided exercises sheets, and practical work sheets.

LANGUAGE

English

CONTACT

Given the central role of machine learning algorithms in many data science tasks, data scientists must be aware of a wide range of machine learning approaches, as well as the field's long history. A data scientist must know where to look for possible techniques to apply to new problems. A data scientist must also be aware of cross cutting concepts, such as the need to evaluate performance and the general classes of challenges encountered in machine learning. This course provides both the overview and the technical knowledge required.

Modifié le : 14 juillet 2026

UE : UE7-DATA

COURSE : Machine Learning

SCOPE OF THE COURSE

OBJECTIVES

COMPETENCE AND SKILLS

- Identify and select the appropriate machine learning algorithm family (supervised, unsupervised, deep learning) for a given problem, justifying this choice using objective criteria
- Implement, train, and evaluate classification and regression models by applying best practices : train/test/validation split, cross-validation, hyperparameter tuning
- Analyze and interpret model performance using appropriate metrics (precision, recall, F1, AUROC, RMSE, R^2) and diagnose bias and variance issues
- Apply unsupervised learning algorithms for dimensionality reduction (PCA, ICA, NMF) and clustering (k-means, GMM, hierarchical clustering) on real-world datasets
- Apply multimedia data analysis methods (images, signals) using suitable machine learning models
- Design, develop, and integrate tools and applications leveraging machine learning models for data valorization

PROGRAM

- **Foundations of Machine Learning :**
  - Overview of paradigms : supervised, unsupervised, reinforcement learning, and deep learning.
  - Symbolic vs. numerical learning ; statistical vs. structural/syntactic approaches.
  - Learning as an optimization problem ; data exploration through machine learning.
  - Cross-cutting challenges : data quality, regularization, generalization, curse of dimensionality.
- **Supervised Learning :**
  - Regression and classification tasks : use cases and selection criteria.
  - Bias/variance trade-off ; model complexity and generalization ability.
  - Evaluation protocols : train/test/validation split, cross-validation, hyperparameter tuning.
  - Classification metrics : precision, recall, F1, AUROC, confusion matrix ; regression metrics : RMSE, MAE, R^2.
  - Fundamental algorithms : linear and logistic regression, KNN, Naive Bayes, decision trees.
  - Extensions and ensemble methods : bagging, boosting, random forests, polynomial features.
  - Model diagnosis : high bias vs. high variance ; feature selection and engineering strategies.

- — Multiclass classification; advanced algorithm : SVM (Support Vector Machine).

— **Unsupervised Learning :**

- — Main tasks : clustering and dimensionality reduction.
- — Clustering algorithms : k-means, hierarchical clustering (connectivity vs. centroid), DBSCAN, Gaussian Mixture Models (GMM).
- — Dimensionality reduction algorithms : PCA (Principal Component Analysis), ICA (Independent Component Analysis), NMF (Non-negative Matrix Factorization).
- — Evaluating unsupervised learning results : silhouette scores, inertia, interpretation of components.

— **Deep Learning — Introduction :**

- — Artificial neural networks : architecture, activation functions, backpropagation.
- — Convolutional networks (CNN) for image analysis; recurrent networks (RNN/LSTM) for signals and time series.
- — Tools and frameworks : TensorFlow, PyTorch, Keras.

— **Practical challenges and implementation :**

- — Data preprocessing : normalization, handling missing values, encoding categorical variables.
- — Machine learning pipelines with scikit-learn.
- — Model interpretability : SHAP, feature importance.
- — Model deployment : prediction APIs, MLOps best practices.

BIBLIOGRAPHY

- — Barra, V., Miclet, L. and Cornuéjols, A. - *Artificial Learning — Concepts and Algorithms, from Bayes and Hume to Deep Learning* - Eyrolles, 4th ed.
- — Aurélien Géron - *Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow* - O'Reilly, 3rd ed., 2022
- — Christopher M. Bishop - *Pattern Recognition and Machine Learning* - Springer, 2006
- — Ian Goodfellow, Yoshua Bengio & Aaron Courville - *Deep Learning* - MIT Press, 2016 : `https://www.deeplearningbook.org`
- — scikit-learn documentation : `https://scikit-learn.org/stable/`
- — TensorFlow / Keras documentation : `https://www.tensorflow.org/api_docs`
- — PyTorch documentation : `https://pytorch.org/docs/stable/`
- — Kaggle — Datasets and ML competitions : `https://www.kaggle.com`
- — Papers With Code — State of the art in ML : `https://paperswithcode.com`

REQUIREMENTS

- — **Probability and Random Variables (S5) :** probability distributions, expectation, variance, law of large numbers — foundations of probabilistic models (Naive Bayes, GMM, regression).

— **Stochastic Processes (S6) :** modeling of time series and random signals — useful for sequential models and deep learning.
— **Data Mining (S6) :** data preprocessing, association rules, introduction to clustering — direct foundation for this course.
— **Expected cross-disciplinary skills :**
    — Intermediate-level Python programming (numpy, pandas, matplotlib)
    — Linear algebra basics (vectors, matrices, eigenvalues) and differential calculus (gradient)
    — Understanding of data structures and basic algorithms
    — Ability to read technical documentation and scientific articles in English