# RCP208 - Apprentissage statistique : modélisation descriptive et introduction aux réseaux de neurones

Mis à jour le 25-09-2026

## Code : RCP208

Unité d'enseignement de type cours

6 crédits

Volume horaire de référence (+/- 10%) : **50 heures**

**Responsabilité nationale :**
EPN05 - Informatique / Michel CRUCIANU

**Contact national :**

EPN05 - Informatique

2 rue Conté
33.1.4A
75003 Paris
01 40 27 22 40
Agnès Lapierre
agnes.lapierre@lecnam.net

## Présentation

### Prérequis

Cet enseignement s'adresse aux auditeurs souhaitant acquérir des connaissances de base sur l'analyse des données, la reconnaissance des formes et la fouille de données (*data mining*).
Prérequis obligatoires : avoir suivi le cycle préparatoire de l'EICNAM ou avoir un niveau équivalent (licence).

### Objectifs pédagogiques

Ce cours donne des éléments de base de l'analyse des données et de la modélisation descriptive, ainsi que des principes à mettre en œuvre pour traiter des applications réelles. Une introduction à la modélisation décisionnelle avec des réseaux de neurones est également présentée. L'analyse des données et la modélisation descriptive aident à comprendre les données empiriques issues de phénomènes naturels, économiques ou socio-culturels. Cette compréhension facilite la mise en œuvre de méthodes performantes de construction de modèles décisionnels.
Les méthodes abordées ont de très nombreuses applications dans des domaines aussi divers que la santé, la climatologie, la sécurité, le marketing, la gestion de la relation client, etc.
L'enseignement adopte une approche pragmatique, les séances de travaux pratiques permettant la mise en œuvre systématique des méthodes présentées.
Les unités d'enseignement RCP209 « Apprentissage statistique : modélisation décisionnelle et apprentissage profond », RCP211 « Intelligence artificielle avancée » et RCP217 « Intelligence artificielle pour des données multimédia » sont des suites recommandées de RCP208.

### Compétences

Analyser des données, mettre en œuvre une modélisation descriptive, mettre en œuvre un modèle décisionnel à base de réseaux de neurones multi-couches.

## Programme

### Contenu

Les thèmes abordés dans les séances de cours et de travaux pratiques (TP) sont :

- Applications, nature des problèmes de modélisation et spécificités des données.
- Analyse des données, réduction linéaire de dimension : méthodes factorielles.
- Classification automatique : k-moyennes, DBSCAN.
- Estimation de densités : noyaux, modèles de mélange.
- Réduction non-linéaire de dimension : t-SNE, UMAP.
- Imputation des données manquantes.
- Sélection de variables.
- Réseaux de neurones multi-couches : capacités d'approximation, descente de gradient, apprentissage et régularisation, représentations internes.

Chaque séance de cours est suivie d'une séance de TP permettant de mettre en œuvre les méthodes présentées.
Les TP sont réalisés en utilisant principalement la plateforme Scikit-learn. Une introduction rapide au langage Python, à NumPy, à Matplotlib et à Scikit-learn est prévue lors des premières séances de TP.

### Modalités de validation

- Examen final

### Description des modalités de validation

Examen ; certaines questions peuvent porter sur les travaux pratiques.

### Bibliographie

| Titre | Auteur(s) |
| --- | --- |
| Reconnaissance des formes : méthodes et applications. Ed. Inter Editions. | A. Belaid, Y. Belaid |
| Réseaux de neurones : méthodologies et applications. Ed. Eyrolles. | G. Dreyfus, et al |
| Méthodes factorielles pour l'analyse des données : méthodes linéaires et extensions non-linéaires. Ed. Hermès. | M. Crucianu, J.-P. Asselin de Beauville, R. Boné |
| Probabilités, analyse des données et statistique. Ed. TECHNIP. | G. Saporta |
| Principles of Data Mining (Adaptive Computation and Machine Learning). Ed. Bradford Book. | D.J. Hand, H. Mannila, P. Smyth |