# Clara CARLIER

Data Scientist | Expert in Machine Learning & Deep Learning

*Passionate about artificial intelligence, I'm keen to put my skills at the service of others by taking an active part in technological advances and building the world of tomorrow.*

**Availability:** Immediate **Mobility:** Paris, Gif-sur-Yvette (±30 km), Île-de-France

clara.carlier@outlook.fr +33 6 78 03 23 23 91190 Gif-sur-Yvette Driving licence

clara-carlier claracarlierdf.github.io

## SKILLS

### Programming languages

- Python : NumPy, Pandas, Scikit Learn, Plotly, TensorFlow, Keras, Matplotlib, SciPy, PySpark, datasets, transformers
- R : dplyr, ggplot2, tidymodels, RMarkdown
- Matlab
- C++

### Methodologies

- Data analysis and visualisation : Python, Tableau
- Advanced statistical modelling: sensitivity analysis, multivariate analysis
- Dimension reduction: classical PCA, sparse PCA, functional PCA
- **Machine Learning:** construction of predictive models (k-NN, RF, SVM, KRR, etc.), classification and regression, supervised and unsupervised learning, cross-validation
- **Deep Learning:** design and training of neural networks (CNN, RNN, LSTM, etc.), generative AI (LLM)
- **Bayesian inference:** ABC methods, SMC samplers
- Image and signal denoising: BM3D
- Optimisation: gradient descent, PGD, MGD, stochastic optimisation
- Time series: (partial) auto-correlation, seasonal and trend decomposition, denoising, prediction and generation
- MLOps: version management (GitHub), function and package testing (pytest), API authoring (Flask, FastAPI)

**Development environments:** Databricks, Jupyter Notebook, Spyder, Anaconda, Visual Studio Code, Azure ML, Apache Spark

### Soft skills

- Complex project management: from design to implementation, respecting time and resource constraints
- Autonomy and good interpersonal skills: working independently while collaborating effectively with multidisciplinary teams
- Scientific communication: presenting complex results clearly to both technical and non-technical audiences (LateX, PowerPoint)
- Documentation and technology watch: keeping abreast of the latest scientific advances and incorporating them into practical projects
- Self-training: curiosity and constant commitment to keeping up to date with new technologies, ability to learn new tools and methodologies quickly

## LANGUAGES

- French: mother tongue
- Spanish: mother tongue
- Englis: C1 - TOEIC 825

## CERTIFICATIONS

- Python Essentials for MLOps, *Duke University, Coursera*
- Generative AI with Large Language Models, *DeepLearning.AI & Amazon Web Services, Coursera*

## PROFESSIONAL EXPERIENCE

May 2021 – Sept. 2024

### Data Scientist

**Renault Group and CREST-ENSAE laboratory, IP Paris**

**PhD topic:** Calibration of complex system models to build the digital twin of the autonomous vehicle

**Technical environnements:** Python, Databricks, Azure, Agiles methods

- Manage a project combining academic and industrial objectives
- Implement and develop an innovative and effective digital solution
- Building neural networks to generate time series
- Monitor scientific and technological developments
- Supervise a student on an end-of-studies placement
- Collaborate with engineers and propose solutions to them
- Write scientific reports and articles for international conferences
- Present results internally and at conferences

June 2019 – Aug. 2019

### Image processing engineer (internship)

**Solid state physics laboratory, CNRS, Orsay**

**Internship topic:** Denoising atomic-scale hyperspectral imaging

**Technical environnements:** Python, Spyder, Jupyter Notebook, Matlab

- Develop a solution for denoising images and signals (BM3D)
- Communicate with physicists to identify and meet their needs

## PROJECTS

### Winning team of the Open Data Hackathon DGFiP 2022

**Technical environnements:** Python, Spyder, GitHub, Tableau, K-Means

- Analyse the coverage of the country by DGFiP structures
- Propose a solution to optimise their distribution
- Implement a Python package available on GitHub
- Communicate the results by video and with Tableau

### Project on earthquakes

**Technical environnements:** R, RStudio, RMarkdown

- Build up a complete and reliable database
- Identify the most appropriate methods for prediction
- Implement a package covering all the work involved
- Write a report and give an oral presentation

### Kaggle competition

**Technical environnements:** R, RStudio, GAM

- Clean and format the database supplied by EDF
- Identify and test effective methods for predicting electricity consumption

## EDUCATION

Sept. 2024

### PhD in applied mathematics

**Renault group and CREST-ENSAE laboratory, IP Paris**

- Industrial PhD (CIFRE)
- Statistics, machine learning, deep learning, Bayesian inference

2020

### Master of Statistics and Machine Learning - Mathematics of randomness

**Université Paris-Saclay, Orsay**

- In collaboration with Polytechnique, ENSAE and Télécom (IP Paris)
- Sophie Germain excellence scholarship winner (FMJH and LMH)
  - Parametric and non-parametric statistics, Bayesian statistics
  - Statistical learning, supervised and unsupervised, regression and classification, dimension reduction, Monte Carlo methods
  - Scientific programming, simulation and prediction projects, data formatting and structuring
  - Probabilistic graphical models, reinforcement and online learning, sequential learning, compressed sensing
  - Convex and non-convex optimisation, operations research
  - Probability, Markov chains, martingales and stochastic processes