

Clara CARLIER

Data Scientist | Experte en Machine Learning, Deep Learning & séries temporelles

Passionnée par l'intelligence artificielle, j'ai à cœur de mettre mes compétences au service des autres en participant activement aux avancées technologiques et ainsi construire le monde de demain.

✉ clara.carlier@outlook.fr

☎ 06 78 03 23 23

📍 91190 Gif-sur-Yvette

🚗 Permis B (véhiculée)

in clara-carlier

🌐 claracarlierdf.github.io

COMPÉTENCES

Langages de programmation

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

Méthodologies

- Analyse et visualisation des données : Python, Tableau
- Modélisation statistique avancée : analyse de sensibilité, analyse multivariée
- Réduction de dimension : PCA classique, sparse, fonctionnelle
- **Machine Learning** : construction de modèles prédictifs (k-NN, RF, SVM, KRR, ...), classification et régression, apprentissage supervisé et non supervisé, validation croisée
- **Deep Learning** : conception et entraînement de réseaux de neurones (CNN et LSTM)
- **Inférence bayésienne** : méthodes ABC, échantillonneurs SMC
- Débruitage d'images et de signaux : BM3D
- Résolution de problème d'optimisation : descente de gradient, PGD, MGD, optimisation stochastique
- Séries temporelles : auto-corrélation (partielle), décomposition en saisonnalité et tendance, débruitage, prédiction et génération
- Gestion des versions : GitHub

Environnements de développement : Databricks, Jupyter Notebook, Spyder, Anaconda

Compétences transversales

- Gestion de projet complexe : de la conception à la mise en œuvre, en respectant les contraintes de temps et de ressources
- Autonomie et bon relationnel : travailler de manière indépendante tout en collaborant efficacement avec des équipes pluridisciplinaires
- Communication scientifique : présenter des résultats complexes de manière claire à un public aussi bien technique que non technique
- Documentation et veille technologique : s'informer des dernières avancées scientifiques et les intégrer dans des projets pratiques
- Auto-formation : curiosité et engagement constant pour se tenir à jour sur les nouvelles technologies, capacité à apprendre rapidement de nouveaux outils et méthodologies

LANGUES

- Français : langue maternelle
- Espagnol : langue maternelle
- Anglais : C1 - TOEIC 825

INFORMATIONS COMPLÉMENTAIRES

- **Disponibilité** : immédiate
- **Mobilité** : Paris, Gif-sur-Yvette (±30 km), Île-de-France

EXPÉRIENCES PROFESSIONNELLES

Mai 2021
Sept. 2024

Data Scientist

Groupe Renault et Laboratoire du CREST (ENSAE, IP Paris)

Sujet de la thèse : Calibration de modèles de systèmes complexes pour la construction du jumeau numérique du véhicule autonome

Environnement technique : Python, Databricks, Azure, méthodes Agiles

- Piloter un projet alliant objectifs académiques et industriels
- Mettre en œuvre et développer une solution numérique innovante et efficace
- Construire des réseaux de neurones génératifs pour la prédiction de séries temporelles
- Mener des veilles scientifiques et technologiques
- Encadrer un étudiant en stage de fin d'études
- Collaborer avec des ingénieurs et leur proposer des solutions
- Rédiger des rapports et articles scientifiques dans des conférences internationales
- Présenter les résultats en interne et en conférences

Juin 2019
Août 2019

Ingénieure traitement de l'image (stage)

Laboratoire de Physique des Solides, CNRS, Orsay

Sujet du stage : Débruitage d'imagerie hyperspectrale à l'échelle atomique

Environnement technique : Python, Spyder, Jupyter Notebook, Matlab

- Développer une solution pour débruiter des images et des signaux
- Communiquer avec des physiciens pour identifier leurs besoins et y répondre

PROJETS

Équipe gagnante du Hackathon Open Data DGFIP 2022

Environnement technique : Python, Spyder, GitHub, Tableau, K-Means

- Analyser la couverture du territoire par les structures DGFIP
- Proposer une solution pour optimiser leur répartition
- Implémenter un package Python disponible sur GitHub
- Communiquer les résultats par vidéo et avec Tableau

Projet sur les séismes

Environnement technique : R, RStudio, RMarkdown

- Constituer une base de données complète et fiable
- Identifier les méthodes les plus adaptées pour la prédiction
- Implémenter un package regroupant l'ensemble du travail
- Rédiger un rapport et réaliser une présentation orale

Compétition kaggle

Environnement technique : R, RStudio, GAM

- Nettoyer et mettre en forme la base de données fournie par EDF
- Identifier et tester des méthodes performantes pour prédire des consommations électriques

FORMATION

Sept. 2024

Doctorat de Mathématiques Appliquées

Groupe Renault et Laboratoire du CREST (ENSAE, IP Paris)

- Thèse industrielle (CIFRE)
- Statistiques, machine learning deep learning, inférence bayésienne

2020

Master de Statistiques et Machine Learning - Mathématiques de l'aléatoire

Université Paris-Saclay, Orsay

- En collaboration avec Polytechnique, l'ENSAE et Télécom (IP Paris)
- Lauréate bourse d'excellence Sophie Germain (FMJH et LMH)
 - **Statistiques** paramétriques et non paramétriques, statistiques bayésiennes
 - **Apprentissage** statistique, supervisé et non-supervisé, régression et classification, réduction de dimension, méthodes de Monte Carlo
 - **Probabilité**, chaînes de Markov, martingales et processus stochastiques
 - **Optimisation** convexe et non convexe, recherche opérationnelle
 - **Programmation** scientifique, projets de simulation et de prédiction, mise en forme et structuration des données
 - Probabilistic graphical models, reinforcement and online learning, apprentissage séquentiel, compressed sensing