

**Pierre Gauthier**

✉ pierrezachariasgauthier@gmail.com

☎ (+33) 6.78.15.86.37

## LIENS

🐙 git/pierrzacharias

in in/pierre-gauthier

## PRESENTATION

Actuellement data scientist au CEA. Je suis ouvert à des propositions pour des projets en science des données. Disponible à partir de mi-Mai 2021. Autonome et Analytique.

## INFORMATIONS

6 rue Montesquieu  
Grenoble, France  
24 ans, Français  
Permis B

## LANGUES

Français  
Anglais (B2)

## LOISIRS

Lecture  
Natation  
Musique (basse)

## EXPÉRIENCES PROFESSIONNELLES

EN POSTE -  
SEP 2019

Data Scientist  
CEA

[ 7 mois ]  
Grenoble, France

Travaux au sein du laboratoire LETI dans le cadre du laboratoire commun avec l'entreprise Diabeloop sur le diabète de type 2.

➤ Travail de biostatisticien sur une étude clinique avec le nettoyage des données médicales, la mise en place d'une base de données, la réalisation de représentations graphiques, la recherche de métriques pertinentes, l'analyse statistique. Présentations et discussions des résultats avec différents acteurs. Participation aux activités de soumission et de publication.

➤ Travail sur le développement d'un simulateur python pour la simulation physiologique de patients diabetiques de type 2 à partir de publications dans le domaine.

Utilisés : Python (pandas, sklearn, scipy), SQL, git, Excel

2019

Data Scientist / Data Engineer  
OCTOPEEK

[ 6 mois ]  
Enghien-les-bains, France

Stage de fin d'étude. Développement de plusieurs outils pour l'enrichissement de données.

➤ Travail sur la partie Data d'une application de prospection géographique pour des professionnels. Nettoyage, mise en forme et stockage des données dans SQL et Elasticsearch. Jointure de données de différentes sources à l'aide d'une distance intertextuelle avec des hyperparamètres optimisés. Implémentation des fonctionnalités géographiques SQL.

➤ Développement d'un outil d'extraction de contenus web modulable (développement avec les bibliothèques python Selenium et BeautifulSoup). Mise en place de stratégies pour résister à la détection de robots web et aux changements d'architectures entre les pages web. L'outil vise ainsi à palier aux mises à jours des sites et aux protections par les changements d'architecture. Déploiement de l'outil pour une utilisation en interne sous forme de micro-services avec une scalabilité à l'aide de Docker, RabbitMQ et Traefik. Gestion des logs avec logstash.

➤ Utilisation d'un modèle de gradient-boosting pour le calcul d'une probabilité qu'un professionnel soit retraité à partir des informations obtenues sur son profil.

➤ Utilisation de l'outil de scraping développé pour enrichir les données d'un client. Nettoyage des données fournies, adaptation de l'outil pour ce besoin spécifique, puis mise en production. Mise en forme des données en sortie pour les livrables.

Utilisés : Python(pandas, sklearn, selenium, beautifulsoup), SQL, Elasticsearch, git, RabbitMQ, Docker, Traefik

2018

Stage en Developpement Web  
MICHELANGE

[ 2 mois ]  
Arnouville, France

Développement web principalement front-end. Travail sur le CSS de plusieurs sites, développement de fonctionnalités en PHP.

Utilisés : HTML, CSS, PHP

## FORMATION

---

### 2016-2019 | Ingénieur Civil des Mines

École d'ingénieur généraliste.

Département Génie Industriel et Mathématiques Appliquées (GIMA).

(Probabilités et statistiques, analyse de données, machine learning, séries temporelles, mathématiques financières, équations différentielles, théorie de l'information). Spécialisé en science des données et en programmation.

**École Nationale Supérieure des Mines de Nancy, France**

### 2014-2016 | Classes préparatoires aux grandes écoles

Filière PCSI / PSI\* (Mathématiques, Physique, Informatique, Sciences de l'Ingénieur).

**Lycée Marcelin Berthelot, Saint-Maur-Des-Fossés, France**

### 2014 | Baccalauréat

Baccalauréat Scientifique.

**Lycée Saint Charles à Athis-Mons, France**

## COMPÉTENCES TECHNIQUES

---

Connaissances en mathématiques appliquées, en science des données et en développement informatique ainsi qu'une expérience sur des projets scientifiques divers.

### Apprentissage - Data Analysis

Fondamentaux en apprentissage automatique, analyse de données et statistiques. Séries temporelles. Algèbre, Optimisation mathématique. Connaissances de base en apprentissage profond avec l'analyse sémantique du langage.

- > **Maîtrise** Python (*sklearn, scipy, ...*), Matlab
- > **Familier** R (*glmnet, forecast, ggplot2, bsts*)
- > **Notions** Pytorch

### Data Processing

Bases de données SQL et noSQL. Connaissance de la suite Elasticsearch avec Logstash, Kibana. Traitement des données avec Python. Déploiement de services avec Docker, RabbitMQ et Traefik. Dashboard avec seaborn.

- > **Maîtrise** Python (pandas, matplotlib, seaborn), SQL
- > **Familier** Elasticsearch, Docker, RabbitMQ, Traefik
- > **Notions** Logstash, Kibana

### Software Engineering

- > **Maîtrise** Git
- > **Familier** C++, HTML/CSS, Unix
- > **Notions** Javascript, Scala, Bash

### Outils

Vim, L<sup>A</sup>T<sub>E</sub>X, Markdown, Adobe Suit, Excel, VBA

### Apprentissage automatique en physique quantique

Inria Grand-Est, France

Prédiction de l'énergie d'atomisation de molécules organiques pour la conception de panneaux solaires.

- > Calcul des descripteurs chimiques des molécules.
- > Utilisation d'un modèle de régression à noyaux.

🔗 : [Quantum Mechanics in a Nutshell](#)

Utilisé : Python (*sklearn, pandas, numpy*)

### Modélisation d'un écoulement chaotique

Inria Grand-Est, France

Projet de recherche pour modéliser un écoulement chaotique par une simulation numérique en prenant en compte le caractère non-homogène des champs magnétiques qui génèrent l'écoulement.

🔗 : [Mélange par un écoulement stationnaire de fluide à faible Reynolds](#)

Utilisé : Matlab

### Sorry ARIMA, but I'm Going Bayesian

Confrontation pour la prédiction des séries temporelles de l'approche fréquentiste usuelle (modèle SARIMA) avec l'approche bayésienne par la méthode bsts de Monte-Carlo par chaînes de Markov.

🔗 : [Sorry ARIMA](#)

Utilisé : R (*glmnet, bsts, ggplot2*)

### EDP start up challenge

Réalisation d'un dashboard dynamique avec seaborn dans le cadre d'un [Startup Challenge Européen](#). Le but du projet était de proposer une stratégie pour le déploiement de nouvelles bornes de chargement pour voitures électriques sur le territoire espagnol.

🔗 : [EDP charging points](#)

Utilisé : Python (*pandas, seaborn*)