

Baptiste Gerbouin

06 41 19 06 50 | bgerbouin@gmail.com | [LinkedIn](#) | [github](#) | [portfolio](#)

Data Scientist | Modélisation statistique | Python – Machine Learning



FORMATION

Université de Bordeaux

Master 2 : CMI ISI - Data Sciences / IA - *Major de Promo*

Obtention: 20 Septembre 2025

Université Libre de Bruxelles - Erasmus

Licence 3 : CMI ISI

Obtention : Juin 2023

Certifications : Deep Learning Specialization (Andrew Ng - Coursera)

Obtention : Décembre 2025

COMPÉTENCES

Data Science & IA : Machine Learning, Deep Learning (CNN, RNN, Transformers), NLP, LLM, Fine-tuning.

Programmation: Python(Pandas, Scikit-Learn, PyTorch, TensorFlow), R, SQL(PostgreSQL), NoSQL(MongoDB), React, TypeScript.

Data Viz & Déploiement : Plotly, Dash, Flask, Git, Docker

Fondamentaux Mathématiques : Probabilités, Statistiques Grande Dimension, Optimisation, Algèbre linéaire.

Soft Skills : Capacité d'analyse, autonomie, travail d'équipe, rigueur scientifique.

EXPÉRIENCES

SUEZ - Data Scientist

6 mois : 2025

- Conception d'une base de données spatio-temporelle scalable (>1.5 To, PostgreSQL) intégrant des flux hétérogènes.
- Automatisation des pipelines ETL et optimisation des requêtes : réduction du temps d'extraction **de 24h à 20 min**.
- Développement de **modèles hybrides** (GR5J + Random Forest) et Feature Engineering sur données haute fréquence pour maximiser le score NSE.
- Co-construction d'un **dashboard** (Plotly Dash) avec une équipe **pluridisciplinaire** (Hydrogéologues) pour valider les modèles et supporter la décision.

CARREFOUR - Défi IA : Prédiction de réachat de produits

Septembre 2024 - Février 2025

- Développement d'un modèle prédictif en machine learning pour recommander des produits sur carrefour.fr, en analysant environ 85 millions de lignes issues de données clients et produits.
- Optimisation des recommandations client via des méthodes de deep learning (LSTM) et machine learning (LightGBM).
- Évaluation des performances avec la métrique Hit Rate @10, visant une amélioration significative du taux de réachat.

UBX : Deep Learning et Classification d'images

Septembre 2024 à Février 2025

- Conception d'un pipeline de classification d'images pour la détection automatisée d'éoliennes sous **PyTorch**.
- Entraînement et comparaison d'architectures : **CNN** (From Scratch) vs **Transfer Learning** (ResNet).
- Mise en place de stratégies de **Data Augmentation** et de pré-traitement pour améliorer la robustesse du modèle.
- Optimisation des hyperparamètres et analyse des métriques de performance pour valider la généralisation.

UBX : Open Data Singapore

Janvier 2024 à Avril 2024

- Exploration des données publiques via tableau de bord interactif sur le logement, les transports et la population.
- Création d'un modèle CatBoost pour estimer en temps réel les prix de revente des logements HDB.
- Pipeline de traitement des données, analyses automatisées, hébergement web (Render), gestion via Git et Gunicorn.

UBX : Monte Carlo Tree Search (Machine Learning)

Septembre 2023 à Janvier 2024

- Développement d'un modèle d'apprentissage par renforcement appliqué au Morpion pour optimiser la prise de décision.
- Conception d'algorithmes d'exploration/exploitation pour améliorer les stratégies de jeu via la méthode MCTS.
- Optimisation algorithmique réduisant le temps de calcul de 30% via l'élagage de l'arbre de recherche.

INFORMATIONS COMPLÉMENTAIRES

Permis B / Français, Anglais (TOEIC B2), Espagnol

Bénévolat : Courses solidaires pour les associations 'Les blouses roses' et 'Chiens guides d'aveugles'

Centres d'intérêts : Cuisine & Badminton