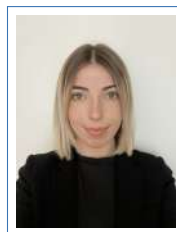


Claire DECHAUX

Data Scientist, Statisticienne et
Ingénieure en Intelligence Artificielle.

📞 06.29.15.71.60
✉ claire.dechaux@ensae.fr
in Claire Dechaux



Expérience professionnelle

- Mai 2024 - **Data Scientist, *Public Health Expertise***, Paris
- Janv. 2025
- Développement d'un outil de Questions/Réponses sur des documents publics de la Haute Autorité de Santé.
 - Modèles et outils utilisés : Création de la base par scrapping de données web, modélisation de type RAG, utilisation de Mistral-7b et Llama3 via Ollama, utilisation de Vanna AI, stockage avec bases de données PostgreSQL et ChromaDB, conteneurisation et déploiement avec Docker, environnement Cloud (AWS).
- Juin 2022 - **Statisticienne - Data Scientist, *Ministère de l'Intérieur***, Paris
- Déc. 2022
- Sections "Enquêtes statistiques" et "Méthodes statistiques et analyses conjoncturelles".
 - Sujet : Modélisation et correction des effets de modes de réponse estimés à partir des données des enquêtes de victimation "Vécu et Ressenti en matière de Sécurité" (VRS) 2022 et "Genre et Sécurité" 2021.
 - Modèles utilisés : Régressions logistiques et modèles d'appariement (matching) sur score de propension.
 - + Missions de NLP sur les données textuelles issues de champs libres des questionnaires d'enquêtes.

Formation

- 2021 - 2024 **Cycle ingénieur, *ENSAE Paris - Institut Polytechnique de Paris***, Palaiseau
- Voie de spécialisation "Data Science, Statistiques et Apprentissage"
 - Statistiques appliquées, Statistiques bayésiennes, Machine Learning, Deep Learning, IA, Traitement du Langage Naturel (NLP), Séries temporelles, Optimisation, Infrastructures et systèmes logiciels, Mise en production des projets de Data Science, Algorithm Design and Analysis, Analysis of Matrix Data, Optimal transport, Ethics in Data Science, Sociologie, Macroéconomie, Microéconomie ...
- Janvier 2023 **Master Informatique et Data Science, *Sapienza Università di Roma***, Italy
- Juin 2023
- Semestre d'échange académique
 - Deep Learning and applied Artificial Intelligence, Big Data computing, Machine Learning, Natural Language Processing.
- 2018 - 2021 **Double-Licence Mathématiques et Economie, *Université Paris-Saclay***, Orsay, Sceaux
- Analyse, Algèbre Linéaire, Probabilités et Statistiques, Traitement de données, Analyse et simulations numériques.
 - Microéconomie, Macroéconomie, Economie publique, Economie internationale, Economie du travail, Sociologie.

Projets et prix académiques

- 2023 - 2024 **Projet d'Advanced ML - "*Predicting molecule-protein interactions for drug discovery*"**
- Prédiction du niveau de pIC50, une mesure de la capacité inhibitrice, de diverses molécules avec une protéine donnée, basée sur leur représentation sous forme d'empreintes digitales ou de graphes.
 - Modèles utilisés : Random Forest, Bayesian Ridge, Graph Neural Networks (GNN)
- Déc. 2023 **Premier prix decerné par Vinci - Hackathon HI Paris**, Institut polytechnique de Paris - HEC
- Tâche : Définir une stratégie, construire des solutions de Machine Learning et d'IA fiables et performantes, trouver un business plan et des solutions à un problème d'efficacité de la supply chain.
- 2023 - 2024 **Projet d'Infrastructures et systèmes logiciels - Développement d'une application**
- Scraping, Python, Java Script, SQL, conteneurisation du Frontend et Backend avec Docker, déploiement.
- 2023 **Projet de Big Data Computing, "*Brain tumor classification with MRI images*"**
- Modèles utilisés : Extracteurs de caractéristiques d'images, Random Forest Model et Multilayer Perceptron Model + Spark
- 2021 - 2022 **Projet de Statistiques appliquées, "*Politique monétaire et prix des actifs*"**
- Projet réalisé pour la Direction générale du Trésor
 - Modèles utilisés : Régressions linéaires (Pooled OLS et Within) et Vecteurs Autorégressifs (VAR)

Langages de programmation

Python, R, SAS, C++, SQL, \LaTeX , Excel + Github

Langues

Français : courant, Anglais : B2 (oral) - C1 (écrit) au *Cambridge Linguaskill certificate*