Clement Guillo

Data Scientist

Status: Head of Statistical Service at INSEE - French Caribbean.
Field: Data Science, Statistical Methods, Spatial Analysis
+33 6 19 93 25 36

Guadeloupe, France
/in/clgui23
github.com/clement2323

Profile

Innovative Data Scientist with a strong foundation in spatial statistics, machine learning, and big data processing. Extensive experience developing and deploying data-driven solutions, including deep learning applications for satellite imagery analysis and automation of complex workflows. Actively seeking opportunities to join data science teams and driving innovation and tackling complex challenges in California's dynamic tech ecosystem.

Professional Experience

Head of Statistical Service - INSEE French Caribbean (Guadeloupe, Martinique, Guiana) 09/2023 - Present

- Developed and applied deep learning algorithms for population census by processing thousands of satellite images of overseas territories (including Réunion and Mayotte). Delivered a trained segmentation model based on these images and deployed the results through an application to support population census estimates.
- Streamlined and automated production practices, coordinating all local teams.
- Managed INSEE agents, overseeing the coordination and execution of on-the-ground survey operations.
- Designed an R training program, aimed at improving the skills of 20 agents.

Head of Spatial Statistics & Data Confidentiality Methods - INSEE - Paris 09/2019 - 08/2023

- Managed a team of 4 and supervised 3 interns specializing in satellite images analysis.
- Developped deep learning models achieving high accuracy in satellite image segmentation for land cover detection.
- Directed the redesign of a georeferencing system impacting 30 million address records.
- Designed and implemented an algorithm to ensure the confidentiality of gridded data across France, addressing differentiation attacks and enabling large-scale data dissemination.
- Supervised statistical disclosure control for 10 public annual INSEE surveys.
- Collaborated with European countries in working groups to harmonize methodologies for spatial analysis and statistical disclosure control at the European level.

Data Science Intern - WESTAT - Rockville MD, USA 07/2019 - 09/2019

- Developed innovative estimation methods for rare events using Zero-Inflated Poisson Models.
- Analyzed large-scale imbalanced datasets.

Household Survey Manager - INSEE - Paris 09/2017 - 08/2018

- Contributed to the redesign of the Master Sample, representing the nationwide household sampling framework for INSEE surveys.
- Conducted large-scale sampling simulations for the entirety of France using advanced sampling methods. The proposed solution, now deployed at INSEE, directly impacts the work of nearly 700 field interviewers.
- Worked closely with a team of five statisticians to parallelize processing workflows and analyze results, including variance estimation

Skills

Programming

- Python (Pytorch, Scipy), R (Expert), SQL (Advanced)
- Docker, kubernetes, MLOPS
- Web Development, Javascript React, NextJS

Languages

- French (Native)
- English (B2) (TOEIC 895/990)

Education

Master of Research in Mathematics, Vision and Learning (MVA) - Polytechnique Paris 09/2019 - 08/2020

- Highest Honors
- Specialized in machine learning and computer vision
- Masters Thesis on statistical models for database fusion
- Deep learning, Reinforcement learning

Statistical Methodology Master - University Rennes 2 - Rennes, France 09/2018 - 08/2020

- Highest Honors
- Advanced coursework in statistical methods and data analysis
- Time series analysis, sampling survey methodology

Engineering Degree in Statistics and Probabilities - ENSAI - Rennes, France 09/2011 - 08/2013

- Core curriculum in statistical methods and programming
- Teaching Assistant for Statistics and Python courses
- Masters Thesis on statistical Methodology for Smoothing Insurance Claims with time series analysis

Projects

Automated Satellite Image Analysis - INSEE - Paris Present

- Developed a full pipeline to detect building locations from Pleiades satellite images.
- Trained SegFormer segmentation models for high-accuracy building detection.
- Integrated a containerized workflow with Kubernetes (Argo Workflows and ArgoCD).
- Built interactive web applications with React (Observable) and GeoServer for result visualization.
- Tracked experiments using MLFlow and hosted the solution on INSEE's Kubernetes cluster.

Automated Study Reporting System - INSEE - Paris Present

- Designed and implemented a comprehensive pipeline for automating study report production using R.
- Developed and published the 'BoutadE' R package on GitHub, enabling rapid, reproducible reporting workflows.
- Automated content generation with local LLMs.
- Integrated advanced visualizations and dashboards.

Hackathon Winner - EY Paris 2018

- Led team of 4 to victory among 20 competing teams.
- Implemented solution using Random Forests in Python.