

Data Scientist

Pablo Laso



Boston, MA (authorized to work in the US and willing to relocate)



929 663 76 99



lasomielgopablo@gmail.com



linkedin.com/in/lasopablo/



github.com/lasopablo



lasopablo.github.io

Skills

Programming

Python, R, C++, Scala, MATLAB, Bash scripting, Linux environments.

Data Science Tools

Pandas, NumPy, scikit-learn, SciPy, Matplotlib, Seaborn, PyTorch.

Data Management & Big Data

SQL (MySQL, PostgreSQL), NoSQL (MongoDB, DynamoDB), Hadoop, PySpark.

Full-stack Development

JavaScript, Node.js, React, CSS, Effect.

Cloud Technologies

AWS (EC2, S3), Azure, Vercel, API deployment (Flask, Django), FastAPI.

Version Control & CI/CD

Git, Docker, Kubernetes.

Data Visualization & BI

Excel, Tableau, PowerBI.

Publications

ISBI 2024 conference (accepted)

Presented with Harvard affiliation. Collaboration with MIT CSAIL lab.

New data augmentation methods, refined priors, and a more robust loss function lead to unprecedented (Dec 2023) performance on a 3D CNN for image segmentation. Link to paper

About

I am a Data Scientist with a robust and diverse technical background in AI and software engineering. Leveraging advanced analytics and machine learning, I specialize in translating complex data into actionable insights that drive strategic business decisions.

Education

MS thesis: AI (full ride. GPA: 4.0), **Harvard University** **Boston, MA**

• Statistics; Deep Learning: Data Augmentation and Algorithm Optimization.

MS Computer Science (GPA: 4.0), **Twente Universiteit** **Netherlands**

• Courses: Design & Analysis of Algorithms, Database Systems, Big Data.

MS Data Science, **KTH - Royal Institute of Technology** **Sweden**

• Courses: Machine Learning, Data Analysis, Business Analysis, Bayesian Inference.

BS Engineering (top national 20 STEM students), **URJC Madrid University** **Spain**

• Courses: Statistics, Programming, Network (DNS, TCP/IP), Signal Processing.

Experience

ML Engineer, Martinos Center (internship) **Boston, MA**

May 2023 - present

- Leveraged PyTorch to mathematically refine the CNN's loss function and optimize data augmentation methods using generative models, outperforming previous state-of-the-art models by 17%.
- Implemented the CNN algorithm into the company's core software using Bash scripting, allowing seamless integration as a new feature accessible via a single command.

Software engineer, MICSII (collaboration) **[hybrid] New York, NY**

Jan 2024-April 2024

- Installation, configuration, and Ubuntu Linux system testing: AWS (EC2/S3) for server/data management, orchestrated with Docker and Kubernetes.
- Enhanced algorithm efficiency by revising I/O functions and integrating advanced features, increasing efficiency by 15% performance and reducing costs in AWS.

Software Engineer, Karolinska Institute (part-time job) **Stockholm, Sweden**

Oct 2021 - Sep 2022

- Managed and optimized server operations on Linux, ensuring high availability and scalability for department-critical operations.
- Administered databases using MySQL and PostgreSQL, focusing on performance tuning and data integrity. Used Bash scripting and Python for automation.

Data Scientist, General Electric (full-time co-op) **Madrid, Spain**

Sep 2021 - Jul 2022

- Implemented robust data handling and analysis pipelines using Python, bash scripting, and SQL, improving the efficiency of data workflows and creating a fully automated system.
- Data visualization for feature analysis and Machine Learning modeling for outcome prediction (from feature processing, feature selection, and dimensionality reduction techniques; to model building and validation) with a 93% accuracy.

Projects

Please explore my projects on my website. You can browse them by categories: Machine Learning | Big Data | Data Analysis | Cloud & Full-Stack Development: https://lasopablo.github.io/vcard_portfolio.