Martín Aramayo

Autonomous City of Buenos Aires, Buenos Aires, Argentina
---------- ---------- m-g-aramayo MartinAramayo MartinAramayo AllMylinks
My blogs: in English blog/Martín, in Spanish blog/Martín/es, (each of the links offers different content).

Completed education

Instituto Balseiro, Full Scholarship, National Commission of Atomic Energy **Jan. 2018 – Dec. 2021**
Master in Statistical Physics (M.Sc.): Applied Simulations & Machine Learning|Physics B.Sc.| Bariloche, Río Negro, Argentina

Experience

Mercado Libre **Dec. 2022 – Oct. 2023 (23 months)**
Data Scientist *Autonomous City of Buenos Aires, Buenos Aires, Argentina*

- Refractor/update of Mod Dev Team Tableau Dashboards (800+ visits), including SQL jobs, redone metrics, and UI/UX
- Implement an image enhancement model to transform 3M+ pictures on the marketplace.
- Launch a masked train feature to our PyTorch CV models, reducing the time requirements for the Data science team to add new moderation conditions from multiple weeks to 0.
- Participation in the internal OpenAI ChatGPT 4 hackathon building a text summarizer tool.
- Tech: OpenAI, ChatGPT, Tableau, Bash, AWS, Big Query, Google Cloud, GAN models, Pandas, Go, SQL, Computer Vision, PyTorch, Python, modelos & BI tools in-house.

Intellignos, Havas **Mar. 2022 – Nov. 2022 (10 months)**
Data Scientist *Autonomous City of Buenos Aires, Buenos Aires, Argentina*

- Improving evaluation metrics efficiency and accuracy for an attribution model converting a costly $O(ne^n)$ metric calculation to a $O(n)$ metric.
- Streamlined ETL analysis, support, documentation, automation, and configuration file design by merging the ETLs of two clients into a single system, reducing the team size and support time required by 50%.
- Reproduced the clients' production dashboard locally with Matplotlib, cutting QA times by half.
- Tech: Bash, DataBricks, StreamSets, BigQuery, Dataproc, PySpark, Azure, Python, in-house models & BI tools.

UNICEF, ONG **Jun. 2020 – Feb. 2022 (1 year 9 months)**
Data Scientist trainee *Bariloche, Río Negro, Argentina*

- Analyzed GIS and microcensus data to identify the five most predictive features for demographic estimation.
- Performed binary classification of occupation status on a group of houses of roughly 100k residents.
- Tech: Scikit-learn, Numpy, Pandas, SPSS datasets, Scipy, SQL.

CNEA, Centro Atómico de Bariloche **Jun. 2018 – Jan. 2022 (3 years 6 months)**
ML trainee — Scholarship Holder — B. Sc. & M. Sc. in Statistical Physics *Bariloche, Río Negro, Argentina*

- Ground-up implementation of agent-based demographic simulation with 300k+ agents. The data pipeline includes: Automatic config & log files, post-execution analysis, testing & benchmarking. Use of computer clusters.
- Mathematical modeling courses: dynamical systems, stochastic models, game theory, technology & healthcare models.
- AI: Machine Learning, Deep Learning, LSTM, NLP, image processing, encoding, feature engineering & clustering.
- Data handling: ETL, data mining from multiple runs of simulation data with Bash and Python scripts.
- Software development: Data Science with Python. Computational models in physics, biology & healthcare projects. CUDA oriented to simulations. Basic data structures in C; trees, stacks, lists.
- Simulated the course correction and orientation control system in a modeled satellite using a Kalman filter.
- Courses on Medical imaging: OpenCV, ImageJ, Matlab, and Python focusing on computational tomographies.
- Biweekly instances of public speaking showing experimental results and technical knowledge acquired.

Projects & skills

Demographic agent-based simulation: Reproduction and resources | *Python, Bash, Pandas, Matplotlib* **April 2021**

- OOP prototype in Python & Final implementation with Pandas reducing in half execution times.
- Public presentations with results and visualizations showing the reproduction of classical models in population dynamics.

Demographic estimation | *Sklearn, Keras, Matplotlib, Python* **August 2020**

- Data exploration and data analysis at the feature engineering stage.
- Implemented binary classifiers to differentiate between inhabited and uninhabited households using satellite data.
- Estimation of subpopulations from an estimated distribution of inhabitants for each household.
- Small dataset estimation, knowing only satellite information and microcensus data.

Programming Languages: Python, C, BASH, SQL
Development tools: Regex, Jupyter, IPython, git, PostgreSQL, BigQuery, Dataproc, Streamsets, PySpark, REST APIS
Technologies/Frameworks: Linux, Docker, Pandas, Sci-kit learn, Keras, Numpy, Matplotlib, PyTorch, Tableau
Documentation: Latex, Markdown, pandoc, Notion, Joplin, Mermaid, Hugo, Inkscape, VS code
Languages: English (ESOL B2), Spanish (Native)