

EXPERIENCE

Mid-Level Data Scientist

Robbyson Systems

May 2024 — Present

São Paulo

- Designed and implemented a Python package for data analysis on GitLab repositories data to assess efficiency and development cost.
- Analyzed data from behavioral projection algorithms in order to ensure quality and consistency.

Tools: Python3, Pandas, Matplotlib, Git, Jupyter Notebook, MongoDB, REST APIs

Junior Data Scientist

Big Data Assessoria Empresarial S/A

Jan. 2022 — Mar. 2024

São Paulo

Joined as an Intern Data Scientist (Jan. 2022 – Oct. 2022), focusing on a client-specific recommendation system. As a Junior Data Scientist, transitioned to the Forecast and Product Operations divisions, where I developed and maintained Machine Learning-centered pipelines.

- Contributed to the development of a recommendation system for a market-leading company in the civil construction sector.
- Idealized and implemented a quality assurance Python framework to automate artifacts validation processes.
- Contributed to code bases with implementation, reviews, maintenance, refactoring and testing.
- Conducted and presented data analyses regarding both client data and the deliverables.
- Developed a price-elasticity modeling approach using Panel Data Regression and Supervised Machine Learning in order to optimize production models.
- Orchestrated ETL and Machine Learning pipelines with Apache Airflow.

Tools: Python3, Pandas, Scikit-Learn, Numpy, Matplotlib, Git, Apache Airflow, Docker, AWS (S3, EC2), PostgreSQL, Jupyter Notebook

Researcher in Applied Machine Learning

São Paulo Research Foundation (FAPESP)

Sep. 2020 — Sep. 2021

Sorocaba

Granted funding for the scientific project proposal (grant #2020/09607-9) in the project “Solar Radiation Forecasting using Machine Learning” for the period of 12 months.

- Conducted and showcased a systematic literature review.
- Elaborated scientific reports regarding the progress of the research, methodology and the results obtained.

Links: <https://bv.fapesp.br/56754>

PROJECTS

Solar Radiation Forecasting using Machine Learning

Sep. 2020 — Sep. 2021

Project regarding use of Machine Learning techniques to obtain models capable of forecasting hourly solar radiation from historical meteorological data 60 minutes in the future. The results obtained were compared to a spatial interpolation technique and an empirical model in order to assess the proposed methods.

- Built a complete data pipeline to pre-process all information from automated download routines to site-specific data normalization.
- Implemented data imputation with Inverse Distance Weighting-based interpolation to artificially reconstruct missing training values.
- Implemented fully automated train/evaluation routines and model selection procedures.
- Configured the laboratory infrastructure to use the Dask distributed processing framework.

Link: github.com/lfenzo/ml-solar-sao-paulo

Tools: Python3, Pandas, Scikit-Learn, Numpy, Optuna, Matplotlib

Impostor.jl – The Highly Versatile Synthetic Data Generator

May 2023 — Present

Impostor is a Julia package which facilitates the generation of synthetic tabular data using a flexible and concise API. Built from scratch upon Julia's *Multiple Dispatch* paradigm with simplicity in mind, Impostor is the concretization of a software engineering project from its data back-end and API design; to its packaging, registration and distribution via the Julia General Registry.

Link: github.com/lfenzo/Impostor.jl

Tools: Julia, DataFrames.jl, Documenter.jl

EDUCATION

Computer Science Bachelor's, Federal University of São Carlos (UFSCar), GPA: 8.93/10

Mar. 2018 — Oct. 2022

SKILLS

Programming: Python3, Julia, C/C++

Cloud: Amazon Web Services (S3, EC2)

Others: Latex, Markdown, Obsidian

Tools: Linux, Git, Matplotlib, Pandas, Numpy, Scikit-Learn, Docker, Apache Airflow, Jupyter Notebooks, Optuna

Languages: Portuguese (Native), English (Fluent), French (Basic), German (Basic)