GUSTAVO DE MARI PEREIRA

+55 (11) 99189-8834 ◇ guhdemari@gmail.com ◇ Sao Paulo, SP, Brazil
linkedin.com/in/gustavodemari ◇ github.com/gustavodemari ◇ gustavodemari.com.br
Languages: Portuguese (Native), English (Proficient), Spanish (Basic)

SUMMARY

M.S. in Computer Science from IME-USP, focused on Reinforcement Learning. Founder of 2 companies, 10+ years of experience working with large-scale databases and building end-to-end ML pipelines. Kaggle competitor and Scikit-learn contributor.

EDUCATION

M.S. Computer Science, *Institute of Mathematics and Statistics*, University of Sao Paulo (IME-USP) Mar 2021 - Sep 2025

Dissertation name: "Reinforcement Learning for Stochastic Shortest Paths with Dead-Ends"
Grants & Awards: Scholarship of Program of Academic Excellence (PROEX) from CAPES
Research topics: Machine Learning, Reinforcement Learning, Probabilistic Planning, Deep Reinforcement Learning
Advisor: Leliane Nunes de Barros

Coursework: Combinatory Optimization, Analysis of Algorithms, Parallel and Distributed Computing, Machine Learning, Artificial Intelligence, Probabilistic Planning, Reinforcement Learning and Reasoning Under Uncertainty

B.S. Computer Engineering, Faculdade de Informática e Administração Paulista (FIAP) Jan 2008 - Dec 2012

Dissertation name: "Neural networks applied to higher education evasion"
Research topics: Machine Learning, Neural Networks, Deep Learning, Open Data, Higher Education, Evasion, Churn reduction
Advisor: Reinaldo Burian

Coursework: Data Structures and Algorithms, Differential and Integral Calculus, Linear Algebra, Artificial Intelligence, Computer Architecture, Operating Systems, Digital Systems and Hardware Design, Computer Networks and Compilers

WORK EXPERIENCE

Founder & Data Scientist, Providentia Jan 2016 - Present

- Engineered a demand forecasting platform using Machine Learning that reduced forecast generation time from days to hours, enabling end-users to make faster and data-driven operational decisions.
- Developed time series forecasting platform that used 5+ models such as exponential smoothing, ARIMA, quantile regression, gradient boosting, recurrent neural networks (LSTM and GRU) using Python/R with Tensorflow/Keras and Pytorch
- Created a back-end and data-pipeline using technologies like Node.js, AngularJS, MongoDB, PostgreSQL, and RabbitMQ, enabling the processing of multiple concurrent forecasts while providing high availability.
- Consulting in Machine Learning and Artificial Intelligence projects (model optimization, chatbots, LLMs and others)

Founder & Software Engineer, Rapi.do Oct 2013 - Dec 2015

- Created a platform to connect motorcycle couriers, businesses, and individuals with Node.js, MongoDB, and AngularJS
- Developed a mobile app for couriers and users, using jQuery Mobile

Data Engineer, Santander Bank Mar 2011 - Jul 2014

- Worked in 30+ statistical modeling and data engineering projects, impacting 10+ million Santander Bank clients
- Created and managed 50+ ETL pipelines using SQL and PL/SQL for databases/tools such as Oracle, SQL Server and SAS
- Migration of manual ETL pipelines from SQL Server to Oracle, reducing the processing time by approximately 83%
- Crafted interactive dashboards and reports using Excel and Pentaho, enabling data analysis and actionable insights
- Pioneered Wiki CRM, enabling analysts to document project details, ETL metadata, and dataset information

Software Engineering Intern, Brazilian Federal Revenue Office Jul 2010 - Mar 2011

- Produced reports for the Brazilian Federal Revenue Office, utilizing BI tools such as Pentaho and Jaspersoft iReport
- Engineered a Java-based communication scheduling tool to streamline inter-departmental interactions
- Constructed administration panels for internal servers, employing HTML, CSS, and Python

SKILLS

Programming: Python, C/C++, Go, R, Julia, Javascript, Java
Data: pandas, numpy, matplotlib, seaborn, SQL, PySpark, Oracle, MySQL, MongoDB, PostgreSQL, pgvector
ML/AI: scikit-learn, xgboost, Keras, Tensorflow, PyTorch, Hugging Face Transformers, LLMs, RAG, Stable-baselines3
Cloud: Amazon Web Services (AWS), Google Cloud Platform (GCP), Microsoft Azure
Techniques: Machine Learning, Deep Learning, Reinforcement Learning, Inverse Reinforcement Learning, Linear Programming
Tools & Misc: Git, Docker, OpenAI Gym, Gymnasium

COMPETITIONS & PERSONAL PROJECTS

- Actively competed in 14 Kaggle machine learning competitions, consistently tackling complex challenges in domains like sales forecasting, finance, and logistics.
- Santander Customer Satisfaction: Developed a classification model to identify dissatisfied bank customers, placing in the **top 30%** of over 5,100 competing teams (Rank 1513/5115).
- Web Traffic Time Series Forecasting: Engineered and tuned predictive models to forecast Wikipedia page traffic, achieving a rank in the **top 32%** of nearly 1,100 teams (Rank 348/1095).
- Impeachment Prediction: Developed a machine learning model to predict legislative votes for the impeachment of a Brazilian ex-president, successfully forecasting the final outcome and achieving high accuracy on the total vote counts.
- Inverse Reinforcement Learning for Epidemic Control: Proposed and implemented an Inverse Reinforcement Learning (IRL) framework using real-world COVID-19 data to learn and validate an optimal reward function for sequential public health policy decisions (e.g., lockdown measures, resource allocation).
- LLM + RAG for Brazilian Companies: Developed a Retrieval-Augmented Generation (RAG) system using Google Gemini Flash and PostgreSQL with pgvector to efficiently answer complex queries about public Brazilian companies, integrating data from the Securities and Exchange Commission of Brazil.
- Equity Risk Premium (Brazil): Performed a time-series analysis comparing the risk-free rate (SELIC/CDI) against equity returns (IBOV) using Central Bank of Brazil (BCB) data. Found the risk-free rate outperformed the Ibovespa across multiple time windows.
- FIFA Team Optimization: Engineered an optimization model using Integer Programming with Python (Pyomo), Julia (JuMP), and solvers like CPLEX/Gurobi to build the highest-rated soccer team under budget and formation constraints.

OPEN SOURCE CONTRIBUTIONS

- Scikit-learn: Contributed to a fundamental machine learning library, focusing on bug fixes, documentation, and improving numerical stability.
- RabbitMQ: Provided maintenance and enhancement for the official documentation.
- TechEmpowerBenchmarks: Modernized the Node.js implementation by updating dependencies and migrating the Express template engine from Jade to Pug.
- MEAN.JS: Upgraded the profile photo upload functionality by integrating a new version of the Multer middleware, including adding tests to ensure code stability.
- Open Knowledge Foundation: Developed a data scraper for the official city gazette of Guarulhos, São Paulo, adding a new source of public data for the project called Querido Diario.

ADDITIONAL INFORMATION

- Conference Volunteering/Attendance: NeurIPS 21, 22, 23, 24; ICLR 22; ICML 21; ICAPS 21
- Academic Contributions: Teaching Assistant of Introduction to Computer Science 2022 (IME-USP); Website for Special Comission of Artificial Intelligence from Brazilian Computing Society