Dr Rafael Garcia-Dias

Contacts & Media

Github

LinkedIn

Google Scholar

London - UK

Skills

Great: 10+ years of experience

- ✅ Python
- ✅ Data Visualisation
- ✅ Data Analytics
- ✅ Academic Research
- ✅ Linux
- ✅ Unix/Bash/Shell

Solid: 5+ years experience

- ✅ Machine Learning
- ✅ Deep Neural Networks
- ✅ Unit Testing, TDD
- ✅ Version control

Good: ~3 years of experience

- ✅ AWS/GCP
- ✅ Docker/Kubernetes
- ✅ MLOps
- ✅ SQL/DBT/Looker
- ✅ R

Academics

- ✅ Coauthor of an ML book
- ✅ 27 Publications
- ✅ More than 4,300 citations
- ✅ h-index: 19, i10: 23

Languages

- ✅ Portuguese - Native
- ✅ English - Fluent
- ✅ Spanish - Fluent
- ✅ Mandarim - Basic

Visa status

No sponsorship needed

Data Scientist

PhD in Astrophysics with more than eight years of experience in Machine Learning applications and more than 12 years of experience in interdisciplinary scientific research.

Senior Machine Learning Engineer

Floe Oral Care

May 2023 - Present

- **Analytical approach**: Responsible for data-driven storytelling, creation of a comprehensive understanding of available data, and architecting the future developments of the product. Including cloud infrastructure architecture, experiment designing and machine learning development.
- **Business Insight**: At Floe, I play a central role. I identify the key problems and challenges we are facing. I design data collection based on critical metrics for the business and conduct analyses that lead to accurate interpretations and valuable recommendations for our customers.
- **Communication**: I craft effective ways of communicating, visualising, and reporting my results and analyses to our internal team. I help prepare investment pitches and scientific outreaches for dental professionals, making our results clear for technical and non-technical people.
- **Technical stack**: I am responsible for the cloud infrastructure, including databases, website setup, and maintenance. I also develop machine learning algorithms for oral health diagnostics. My role involves designing user-friendly data visualisations and implementing both the back-end and front-end of an app for accessing oral health test results. As the sole full-time code developer in the company, I handle a diverse range of tasks. I am also working on developing the company's data architecture for machine learning development.

Decision scientist (Data Scientist at credit risk)

Monzo Bank Ltd

August 2022 - February 2023

- Developing, monitoring and maintaining multiple ML models in production, improving the creditworthiness evaluation process and helping Monzo provide better prices that could potentially impact the over 7 million customers at Monzo, with a lending portfolio of £259M.
- Created and delivered a **GBM**-based model that substantially improved performance compared to previous models (20-100% Gini gains) in the evaluation of overdraft applications using **Docker** containers and the **Google Cloud Platform.**
- Restructured the internal library to improve test coverage and standardise monitoring across Monzo products. This led to valuable insights into the performance of the models on various subgroups of customers that have been previously ignored.
- Data management and analysis using **BigQuery**, **SQL**, and **DBT** data models. Help stakeholders make decisions by summarising and presenting insights from the data analysis using **Jupyter notebooks** and **Looker** boards.

Postdoctoral Research Associate

🧠King's College London

August 2018 - August 2022

- 🎯 Technical leadership of the Neurofind.ai project, a web-based tool to aid the diagnosis of mental disorders using 3D images of the brain.
  - Developed Neurofinf.ai kernel model using tools like **MLFlow**, **Scikit-learn** and **TensorFlow**. Designed and deployed the containerised backend of Neurofind.ai with **Doker** and built an API using **Flask**, **Celery**, **Flower**, and **Redis** to communicate the ML backend with the website's front end.
  - Designed and implemented the visualisations in the final Neurofind.ai product using **Matplotlib**, **Pandas** and **Numpy**.
  - Data cleaning and preprocessing of more than 100 thousand images.
- 🎯 Created Neuroharmony, an ML tool aimed at mitigating bias across different scanners. This was a major contribution to bridging the gap between academic research and clinical implementation of ML. The work was published in NeuroImage, the most important journal field. The article received 41 citations so far, including mentions in Nature and resulted in a £109,000 MRC research grant. The project was selected among 195 applications.
- 🎯 Authored a chapter and co-authored five additional chapters in the book "Machine Learning: Methods and Applications to Brain Disorders", among other high-impact publications.

PhD Fellow on Machine Learning Applied to Astrophysics

🔭Instituto de Astrofísica de Canarias (IAC) - ULL

August 2015 - August 2018

- 🎯 Applied ML techniques to analyse a large, high-resolution spectroscopy dataset of over 250,000 stars within our galaxy across more than 8 thousand features.
  - Explored **unsupervised learning** algorithms to group stars based on their spectroscopic properties. The study resulted in one of the first ML publications at the A & A Journal, which is the main publisher of the subject.
  - Explored **dimensionality reduction** and **supervised learning** algorithms to trace the origins of stars across various stellar clusters. In this work, I pushed the limits of chemical tagging, an important concept in Glactic Astrophysics. The work was also published in the A & A Journal.
  - Awarded with a PhD Cum Laude, the highest possible award for a PhD at ULL.
  - Gained extensive experience in applying ML to complex scientific datasets, which improved my skills in data analysis, algorithm development, and scientific communication.

Master's Degree in Physics (fully funded)

⚛Universidade Federal do Rio Grande do Sul (UFRGS)

August 2013 - August 2015

- 🎯 Researched the chemical composition and ages of stellar clusters using a combination of data analysis and **Python** programming.
  - Created and implemented a Python program to perform a meta-analysis of dozens of articles from the literature and extracted data to reanalyse over 60 stellar clusters in a homogeneous and automated manner. Gained valuable experience in developing complex programs in **Python**, as well as learning **Fortran** and **Bash scripts**, **version control**, **unit testing**, **data visualisation**, and **data analysis**.

Bachelor of Science Degree in Physics

⚛Universidade Federal do Rio Grande do Sul (UFRGS)

August 2008 - August 2013