

Dr Rafael Dias

PhD in Astrophysics with more than 7 years of experience developing Machine Learning applications and more than 10 years of experience in professional Python programming.

✉ rafaelagd@gmail.com  [GitHub](#)  [LinkedIn](#)  [Google Scholar](#)  [London - UK](#)

Visa status: Right to work in the UK - **No sponsorship needed**

Portfolio of product creation and implementation:

 [Neurofind](#)  [Floe](#)  [ML Book Chapter](#)  [ML article 1](#)  [ML article 2](#)  [ML article 3](#)

Data Scientist / Machine Learning Engineer

 **Floe Oral Care - London - UK**

■ **May 2023 - Present**

Tech Stack: [AWS](#) [CI/CD](#) [Django](#) [Docker](#) [Machine Learning](#) [MLOps](#) [MySQL](#) [Plotly](#) [Python](#) [Scikit-Learn](#)

- 🎯 **Analytical approach:** Responsible for data-driven storytelling, creating 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 fundamental 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. I set up and maintained the CI/CD platform, the cloud services and the integration of the various data sources to guarantee smooth logistics and a clean platform for development and constant improvement.

Data Scientist at Credit Risk (Decision scientist)

 **Monzo Bank Ltd - London - UK**

■ **6 Months - August 2022 - February 2023**

Tech Stack: [BigQuery](#) [Continuous Integration](#) [dbt](#) [Docker](#) [GCP](#) [Looker](#) [Scikit-Learn](#) [Machine Learning](#) [MLOps](#) [Parquet](#) [Plotly](#) [Python](#)

- 🎯 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.
- 🎯 Developed and delivered a **GBM**-based model that substantially improved performance compared to previous models (20-100% Gini gains) in evaluating 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 on **Jupyter notebooks**, **Presentations** and **Looker** boards.

Data Scientist

 **Neurofind / King's College London - London - UK**

4 years - August 2018 - August 2022

Tech Stack: **AWS** ▾ **Continuous Integration** ▾ **Deep-Learning** ▾ **Docker** ▾ **Kubernetes** ▾
MLFlow ▾ **MLOps** ▾ **Plotly** ▾ **Python** ▾ **R** ▾ **Scikit-Learn** ▾ **Tensorflow** ▾

- Developed the backend, cloud infrastructure and ML models for [Neurofind.ai](#), a tool to aid the diagnosis of mental disorders using 3D brain images.
- Published the first out-of-sample harmonisation tool for neuroimaging, [Neuroharmony](#), an ML tool to mitigate bias across different scanners, a significant contribution to bridging the gap between academic research and clinical implementation of ML.
- Collected a £109,000 [MRC grant](#).
- 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 - Tenerife - Spain**

3 years - August 2015 - August 2018

Tech Stack: **Git** ▾ **Linux** ▾ **Machine Learning** ▾ **Matplotlib** ▾ **Numpy** ▾ **Pandas** ▾ **Python** ▾
Scikit-Learn ▾

- Created one of the first applications of ML techniques in astrophysics, analysing a high-resolution spectroscopy dataset of over 250,000 stars within our galaxy across more than 8 thousand features.
- Explored **unsupervised learning** algorithms on the innovation of the stellar classification field.
- Explored **dimensionality reduction** and **supervised learning** algorithms to trace the origins of stars across various stellar clusters.
- Awarded with a PhD Cum Laude, the highest possible award for a PhD at ULL.

Master's Degree in Physics

 **Universidade Federal do Rio Grande do Sul (UFRGS) - Brazil**

August 2013 - August 2015

Tech Stack: **Git** ▾ **Linux** ▾ **Math** ▾ **Matplotlib** ▾ **Numpy** ▾ **Pandas** ▾ **Python** ▾ **R** ▾
Statistics ▾

- Researched the chemical composition and ages of stellar clusters using data analysis and **Python** programming.
- Developed a tool to perform a meta-analysis of over 60 stellar clusters in a homogeneous and automated manner.

Bachelor of Science Degree in Physics

 **Universidade Federal do Rio Grande do Sul (UFRGS) - Brazil**

August 2008 - August 2013

Main topics: **Differential Equations** ▾ **Linear Algebra** ▾ **Math** ▾ **Physics** ▾ **R** ▾ **Statistics** ▾

Proficient Languages: English, Spanish, Portuguese