

# Rafael Garcia-Dias

## Data Scientist

PhD in Astrophysics with more than 7 years of experience in ML applications and more than 12 years of experience in interdisciplinary scientific research.

## Contacts & Media

 [rafaelagd@gmail.com](mailto:rafaelagd@gmail.com)

 [@garciadias](https://twitter.com/garciadias)

 [@garci-dias](https://www.linkedin.com/company/garci-dias)

 [Rafael Garcia-Dias](https://plus.google.com/+RafaelGarciaDias)

 Romford, 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 - git/GitHub/bitbucket

**Good: ~2 years of experience**

- AWS/GCP
- Docker/Kubernetes
- Containerised Applications
- SQL/DBT/Looker
- R

## Academic records

- Coauthor of an ML book
- 26 Publications
- More than 3,800 citations
- h-index: 18, i10: 23

## Languages

- Portuguese - Native
- English - Fluent
- Spanish - Fluent

## Visa status

Right to work - No sponsorship needed

## 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 Neurofind.ai kernel model using tools like **MLFlow**, **Scikit-learn** and **TensorFlow**. Designed and deployed the containerized backend of Neurofind.ai with **Docker** and built an API using **Flask**, **Celery**, **Flower**, and **Redis** to communicate the ML backend with the website's frontend.
  - 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 on 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 analyze 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 Galactic 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 reanalyze over 60 stellar clusters in a homogeneous and automated manner. Before my work, these objects were analyzed using a range of methods that relied on extensive human interaction and subjective interpretation.
  - Gained valuable experience in developing complex programs in **Python**, as well as learning **Fortran** and **Bash** scripts, **version control**, **unit testing**, **data visualization**, and **data analysis**.

## Bachelor of Science Degree in Physics

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

August 2008 - August 2013

## Internships

*Astrophysics Laboratory - UFRGS*

August 2011 - August 2013

- Developed **Python** and **Bash** script algorithms for Astrophysical image processing. The algorithms improved the photometric detection of stars by 2 orders of magnitude. The work was presented at an international congress of Astrophysics and inspired the improvement of photometry on astrophysics surveys.
- Developed C algorithms to simulate the dynamics of star cluster disintegration.

*Magnetism Laboratory - UFRGS*

August 2010 - August 2011

- Developed experiments in nanotechnology related to Giant Magnetoresistance

## Recorded Public Talks

Talk at the PPGFSC physics seminars 📺 📺

2021 - UFSC, Florianópolis, Brazil

Machine learning: potential and limitations

Talk at the Big Data London 2019 📺 📺

2019 - London, United Kingdom

Why do some machine learning models fail?