# FRANCISCO ALMEIDA

DATA SCIENCE BIOINFORMATICS

### BIO

I'm a scientist with a strong interest in the convergence of science, technology, and data. Excited about unraveling molecular mysteries and leveraging programming skills for transformative applications using data science and machine learning. Presently, I took some time from business to focus on learning about specific topics. I'm motivated to take on new challenges and make meaningful contributions.



# **CONTACT**

- ☑ franciscoal95@hotmail.com
- in linkedin.com/in/fransci-co
- github.com/fransci-co
- fransci-co.github.io

# **WORK EXPERIENCE**

### **Data Scientist**

@ Kantar Worldpanel - CoE, Lisbon - May 2023 - October 2024 (1 yr 7 mos)

- Redesigned legacy products originally developed in SAS and transitioning to R, achieving runtime reduction and improved computational efficiency. Implementation of R best practices for development and package creation.
- Developed and managed the ETL data pipeline for the Central and South American regions to power the global advanced analytics data products.
- Delivered and maintained advanced analytics reports to some of the largest global clients. Consumer panel clustering, segmentation, classification and other solutions across time-series.

### **Data Analyst**

@ Kantar Worldpanel - CoE, Lisbon - April 2022 - May 2023 (1 yr 2 mos)

- · Performed statistical analysis and modeling on consumer panel data to uncover insights that drive brand growth.
- Delivered customer, brand, advertising, media and communication analytics, extracting key performance indicators (KPIs) to enhance brand strategies and foster innovation.
- Worked closely with stakeholders from all over the world to ensure the insights and KPIs obtained suited their needs.

#### **Bioinformatician - Trainee**

@i3S - Instituto de Investigação e Inovação em Saúde, Porto - Hybrid / 2020 - 2022 (1y 4 mos)

- Next Generation Sequencing data processing and analysis pipeline development. Variant Calling from Whole Exome Sequencing and Whole Genome Sequencing data with clinical applications using multiple variant callers.
- Run pipelines on high performance computing environment with unix system.

### **EDUCATION**

### MS, Bioinformatics & Computational Biology

2019 - 2021

University of Porto

Master Thesis:"Analysis and comparison tool for multiple variant calling results from germline NGS data with clinical applications"

i3S, Porto, Portugal - Expression Regulation in Cancer group.

#### BS, Genetics and Biotechnology

2016 - 2019

University of Trás-os-Montes e Alto Douro

### **MY STACK**

### TECH

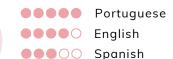
</> python, R, SQL, SAS, matlab

Shiny, plotly

Git, Docker, Bash

Azure, Databricks

### LANGUAGES



### .

#### **DATA & MACHINE LEARNING**

TensorFlow		arrow	, purrr	tidymodels		PyTorch	n Scik	Scikit-learn	
caret	ggplot2		statsmodel	dplyr	da	ta.table	pandas	NumPy	

## **PUBLICATIONS & COURSEWORK**

- Co-author in publications in international journals [1,2].
- Multiple panel and oral communications in national and international conferences in the fields of Genetics and Bioinformatics Here

#### **REFERENCES**

1. https://doi:10.1016/j.jgar.2019.05.017 2. https://doi.org/10.1007/s10096-019-03709-6

### **Licenses & Certifications**

#### Coursera

Python Essentials for MLOps (Credential ID: <u>3L7CJZVICL0U</u>)
Exploratory Data Analysis for Machine Learning (Credential ID: <u>WZNM54SUKPEC</u>)
Machine Learning with Python (Credential ID: <u>XSCWRJX8FU3W</u>)

#### Academic

(Valued with 6 ECTS)

Applied Statistics \ Introduction to Data Science \ Programming and Databases \ Image Processing and Analysis \ Advanced Topics in Data Science \ Computer Vision \ Data Visualization \ Data Mining \ Algorithms for Bioninformatics \ Bionformatics for Omics \ Data Structures for Bioinformatics \ Next Generation Sequencing \ Computational Biochemistry

### **LATEST PROJECTS**

Experience in various ML algorithms both in supervised and unsupervised learning

- Random-forests and XGBoost for binary classification and model explainability packages such as DALEX for model interpretation. Variable importance ranking with model agnostic methods.
- Logistic regression to uncover and measure the relationship between a set of variables and specific KPIs.
- Model optimization with different methods, regularization, hyperparameter tuning, weighting.
- Clustering and dimensionality reduction algorithms e.g. (hierarchical clustering/PCA).
- Experience with Deep Learning models for different applications, e.g. in context of computer vision- object segmentation and classification.
- Web applications using shiny & streamlit.