# Rosamaria **Carraro**

PhD in Astrophysics · DataCamp certified Professional Data Scientist

carrarorosamaria@gmail.com | rosamariacarraro.ml | GitHub | BitBucket | LinkedIn | Google Scholar

## Summary

I am a Ph.D. in Astrophysics with international experience. My great analytics skills and thirst for learning are complemented by my effective communication abilities, showcased through talks at international conferences and published papers. I am eager to apply my expertise and effectively collaborate and communicate with diverse teams and stakeholders leveraging my strong language skills in English, Italian, and Spanish.

## Computer skills

| | |
|---|---|
| **Programming** | Python (pandas, sklearn, multiprocessing), SQL, Git, Unix shell. |
| **Machine Learning** | Classifications, Regressions, Dimensionality reduction, Ensemble learning, Random Forests, Boosting, Model tuning. |
| **Deep Learning** | PyTorch, TensorFlow, NLP, Convolutional Neural Networks, Recurrent Neural Networks, Attention models. |
| **Data Visualization** | Matplotlib, Seaborn, Bokeh, Looker, Keynote, LaTeX. |
| **Mathematics** | Calculus, Statistics, Linear Algebra. |
| **Others** | Regular Expressions, Monte Carlo methods: bootstrapping. |

## Portfolio

NLP on hair products ingredients [link] In this project, I study ingredient lists of a hair product line to understand their properties, by translating differences between their ingredients to distances on a plane. This involves web scraping, One-Hot encoding, and t-SNE dimensionality reduction. I use an interactive Bokeh app for easy product distance visualization and ingredients comparison.

Data Challenge using K-NN [link] An ML classification project using a Kaggle dataset following all the basic steps, from exploratory data analysis to model evaluation. I used Principal Component Analysis for dimensionality reduction, K-Nearest Neighbors, and many visualizations.

Light curves [link] A research project studying the variation of active black holes at the centers of galaxies. I extracted light curves (time series) of variable sources from their raw images. I cross-correlated light curves with different physical origins and I simulated them, in order to retrieve additional constraints on their physical properties.

A Semi-Empirical model for galaxies [link] A research project in which we generated mock catalogs of galaxies and studied the drivers of the statistical relations among galaxy physical properties found in the literature.

Pizza recipe generator [link] An interactive Bokeh app that delivers a pizza recipe based on your tray size and favorite thickness.

Django workout calendar [link] A web app that allows you to keep track of the workouts you perform by following YouTube videos.

Co-evolution of black holes and host galaxies [link] My first published work, studying a large sample of galaxies selected from flat tables: I cross-matched the tables and selected sources based on their properties, to finally extract meaningful statistics and confidence ranges for my samples of sources.

## Certifications

| | | |
|---|---|---|
| DeepLearning.AI | **Deep Learning** Specialization on Coursera | *2022* |
| DataCamp | **Professional Data Scientist** Certification | *2022* |
| DataCamp | **SQL Fundamentals** Skill Track | *2022* |
| DataCamp | **Data Scientist with Python** Career Track | *2020* |
| DataCamp | **Statistics fundamentals with Python** Skill Track | *2020* |

## Education

**PhD in Astrophysics** — *Valparaíso, Chile*

Universidad de Valparaíso — *2015 - 2022*

Yale University - Visiting Assistant in Research — *Jan – Nov 2017*

**Master's Degree in Astronomy** — *Padua, Italy*

Università degli Studi di Padova, Final grade 110/110 — *2010–2015*

**Bachelor's Degree in Physics** — *Rome, Italy*

Università degli Studi Roma Tre, Final grade 104/110 — *2004-2010*