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, and I am eager to bring the extensive data analysis skills I built during my PhD to new challenges. I possess strong analytics and effective communication abilities showcased through talks at international conferences and published papers. With a constant thirst for learning, I am excited to collaborate with diverse teams leveraging my fluency 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.

Portfolio

SENTIMENT ANALYSIS ON USER COMMENTS [LINK]   Performed multi-label text classification on 5,500 user comments to classify 26 emotions. Utilized TF-IDF vectorization, data cleaning, and stemming. Explored independent binary classifiers and classifier chain approach. Developed a KPI indicating 39% positive sentiment in the sample of comments.

NLP ON HAIR PRODUCTS [LINK]   Built a tool to estimate hair product similarity based on web-scraped data using one-hot encoding and t-SNE dimensionality reduction. I developed and hosted an interactive Bokeh app to visualize product distance and compare ingredients.

DATA CHALLENGE USING K-NN [LINK]   An ML classification project using a Kaggle dataset consisting of 118 anonymized features. I used Principal Component Analysis (PCA) for dimensionality reduction, K-Nearest Neighbors (KNN), and many visualizations.

LIGHT CURVES [LINK]   PhD research project investigating the variability of active black holes in galactic centers. Extracted light curves (time series) from raw images of variable sources. Employed cross-correlation techniques to study light curves with diverse physical origins. Simulated light curves to obtain additional constraints on the physical properties of the sources.

A SEMI-EMPIRICAL MODEL FOR GALAXIES [LINK]   PhD research project involving the generation of mock catalogs to analyze the drivers of statistical relations among galaxy physical properties. Validated existing literature findings, advancing our understanding of galaxy properties.

PIZZA RECIPE GENERATOR [LINK]   Developed and hosted an interactive Bokeh app that generates personalized pizza recipes based on user-selected tray size and thickness.

DJANGO WORKOUT CALENDAR [LINK]   Created a web application that enables users to log and track their workouts performed by following YouTube videos. Implemented features for workout logging, and video integration, providing a seamless experience for tracking fitness routines.

CO-EVOLUTION OF BLACK HOLES AND HOST GALAXIES [LINK]   PhD research project 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*