

Tommaso Derossi

Data Scientist — Remote (10–20h/week)

✉ Via I. Campioni 20, Langhirano (PR), Italy | ☎ +39 3486612907 | 📩 tommasoderossi7@gmail.com | 🔍 LinkedIn | 🔍 Github | 🔍 Personal Website

Summary

Data scientist and ML engineer experienced in collaborating with the most diverse stakeholders. I build decision-support systems end-to-end (ETL → modeling → evaluation → dashboards) and translate messy data into evidence-based, actionable insights.

Domains: Research | Government | NGOs | Enterprises | Startups

Stack: Docker | AWS (S3/EC2/Lambda) | Spark | Python [Pytorch - Fastapi - Flask - MLflow - Metaflow - W&B - Airflow] | R | Java | Javascript

Experience

Supervised Program for Alignment Research

September 2025 - Present

Remote

Research Fellow

🔗 <https://sparai.org/>

- **Alignment research** with SPAR: CoT monitoring, decomposition techniques, and causal attribution methods to analyze LLM reasoning traces.

GEA Procomac SPA

November 2024 - Present

Industrial PhD

🔗 <https://www.gea.com/en/>

- Designing and developing an embedding-based **recommendation system** for troubleshooting industrial plant machinery (patented project).

Freelance

September 2024 - Present

Data Scientist / ML Engineer

Remote

- Developing a skills-based **job–candidate matching** system with a focus on explainability and fairness.

Tomorrow S.R.L.

February 2023 - July 2024

Data Scientist

🔗 <https://www.tomorrowsrl.com/>

- Conceptualized and developed data transformation pipelines integrated with **machine learning** models (Bayesian hierarchical models, Random forest, Gradient boosting) for the analysis and presentation of **IOT data** (Predictive maintenance, KPIs optimization, Data-driven actionable insights, Time series forecasting)
- Designed an AI conversational agent for purchasing assistance in VR spaces based on **LLMs, speech-to-text and text-to-speech** (prompting/fine-tuning, RAG, LlamaIndex, LangChain, ElevenLabs).
- Built a **recommendation system** and visualization dashboard for data-driven soccer scouting decision-making support, based on massive **match-events data** and **machine learning** models (Pytorch, Sklearn)
- Created **marketing automation pipelines** from **target audience selection** and **content production** to **campaign creation and management** with **human-in-the-loop** for different advertisement platforms (Linkedin, Meta, X)

WATERONMARS S.R.L.

March 2022 - February 2023

Data Scientist

🔗 <https://www.waternmars.it/>

- Built **ETL pipelines** (web scraping, APIs) and curated analysis of web content, in particular, **social media data** (textual data, images, videos, geospatial and temporal data)
- Implemented and integrated **predictive models** for data enrichment in the domains of **computer vision** (OpenCV, SimpleCV) and **natural language processing** (Spacy, Transformers, NLTK, Gensim, TextBlob, Prodigy)
- Created highly customized **data visualization dashboards** for various clients and types of analysis (Geospatial analytics - GIS, Election campaign analysis, Brand monitoring, Buyer persona, USP analysis, E-commerce analytics)

OMIGRADE S.R.L.

March 2019 - October 2020

Machine Learning engineer

🔗 <https://www.omigrade.it/>

- Designed, prototyped, and optimized a model based on **deep recurrent neural networks** (LSTM, Tensorflow) for the **interactive classification of relational databases' columns**
- Engineered and integrated in a **distributed environment** (Hadoop cluster) the machine learning-based model prototyped (**MLOps**: standardized data preparation, Distributed data with spark's RDD, automated periodic distributed re-training, monitoring, and performance analysis) + Exploration of model variations and data post-processing techniques

Performance: 60 classes, Top-1 accuracy: $89.0 \pm 1.5\%$, Top-5 accuracy: $94.1 \pm 0.3\%$

Education

Università degli studi di Parma

November 2024 - Present

PhD Student

Deep Learning

🔗 <https://www.unipr.it/>

- Currently doing research on Artificial Neural Networks **safety and interpretability** with a focus on **Large Language Models** working on combinations of **bottom-up (mechanistic interpretability)** and **top-down (behavioral experiments)** approaches.

Università degli studi di Torino

September 2019 - April 2023

Master Degree

Stochastics and Data Science

🔗 <https://www.unito.it/>

- Degree course in **Mathematics** fully taught in **English**

Thesis: "Modelling the discourse around nature in Twitter in presence of highly imbalanced data."

Thesis topics: Natural Language Processing, Transformers models, Network Science, Filtering Techniques, Sentiment Classification, Semantic Clustering, Topic Modeling

KU Leuven

September 2020 - July 2021

1 year exchange programme

Statistics

🔗 <https://www.kuleuven.be/>

- Second year of master's degree spent in Belgium at KU Leuven. Some of the main courses attended (10 courses in total): **Optimization and Numerical Methods** (17/20), **Concepts of Bayesian Data Analysis** (16/20), **Data Mining and Neural Networks** (15/20), and **Modern Data Analytics** (15/20)

- Participated in **Data4Good Challenge (Emergent Leuven)**

Università degli studi di Parma

September 2016 - September 2019

Bachelor Degree

Physics

🔗 <https://www.unipr.it/>

- Awarded for being the **top-ranked student in credits earned relative to grade point average** during the second academic year

Thesis: "Methodological aspects of neural networks integrated within an application for the classification of relational database columns."

Thesis topics: Artificial Neural Networks, Deep Recurrent Neural Networks (LSTM), Distributed Model Training

Liceo Scientifico "Giacomo Ulivi"

2011 - 2016

Scientific Oriented

High School

South African College High School

2012

English School Attendance, 3 months

17

Skills

Python

R

Java, Javascript

MATLAB

SQL, NoSQL

MS Office (Excel), LaTeX

ML/DL (Pytorch, Tensorflow)

Spark, PySpark

Visokio Omniscope

Docker

AWS (S3, EC2, Lambda)

MLflow, Weights & Biases

Metaflow, Airflow

Modeling

Forecasting

Dashboarding

Languages

Italian

Mother tongue

English

Fluent

Spanish

Intermediate