Sebastiano Ferraris

sebastiano.ferraris@gmail.com • London, UK • British and Italian citizenship

[Linkedin](#) • [Github](#) • [Blog](#) • [Google Scholar](#) • [Research Gate](#)

Data Scientist and Researcher 10+ years experience in researching and developing analysis, prototypes, and algorithms from proof of concept to production. Proven track records of implementing, validating and scaling algorithms to solve a range of research and industrial problems. Keen on addressing the challenges posed by AI and big data systems around productionisation, deployment and algorithms continuous validation. Scientific author published in international journals.

Experience

JUNE 2024 – TODAY
Senior Data Scientist | [Kpler](#) | London, UK
Leading provider of technology-led data, analytics, and market insights focused on the energy and shipping markets.

- Developing prototypes and algorithms to automate AIS data analysis.
- Collaborating with Analysts to provide fast solutions to their specific needs.
- Collaborating with Dev teams to productionise proof of concepts into production.

JUNE 2020 – JUNE 2024
(LEFT DUE COMPANY INSOLVENCY)
Data Scientist | [General System](#) | London, UK
Startup in stealth model until 2022. High performant real-time analytic platform for high volume (100+Bn) spatiotemporal data

- Designed and wrote production code for a novel, robust and linear-time clustering algorithm to detect dwells in mobility data in **Python, scikit-learn, pandas, numpy, streamlit, DeckGl, KeplerGl**.
- Developed a hierarchical density based algorithm prototype for spatiotemporal data.
- Created on-line and batch outlier detections and corrections algorithms for spatiotemporal data.
- Researched and prototyped two linear-time data fusion algorithms for detecting co-locations across multiple layers, such as AdTech, AIS and ADS-B datasets.
- Leveraged these algorithms to detect: dark vessels, crowds gathering, consumers' patterns, and cross visitations, setting up data processing pipelines with **Databricks** and the internally developed **Data Flow Index**.
- Worked closely with the Front and Back End production teams to turn prototypes into scaled up products, with **AGILE** and **twelve factor app** methodology, with **CI/CD, unittesting, integration testing, contract testing**.
- Won the first internal hackathon with a project on embedding satellite images into the General System's spatiotemporal platform.
- Supported customer success and marketing supporting the creation of visualizations and presentation materials.
- Open sourced [a python library and a series of examples](#) for analysts to interact with the General System's spatiotemporal platform.

- Contributed to the [company blog](#).
- Presented prototypes and findings to stakeholders and potential clients.

SEPTEMBER 2019 – JUNE 2020

(LEFT DUE TO COVID-19 DISRUPTION IN THE HOSPITALITY INDUSTRY)

Algorithm Engineer | Pace revenue management (now [flyr](#)) | London, UK

Startup providing predictive analytics and dynamic pricing for the hospitality industry in a cloud solution integrated in the PMS

- Part of the simulation and validation team.
- Participated in developing an **agent based simulation** aimed at validating the core prediction algorithm.
- Maintained **Python** and **SQLAlchemy** production code with the Back End team.
- Migrated production codebase from **Pandas** to **Dask** to improve scalability..

OCTOBER 2018 – JUNE 2019

(LEFT TO PURSUE A CAREER IN DATA SCIENCE MORE ALIGNED WITH MY STUDIES)

Back end developer | [Thought Machine](#) | London, UK

Cloud-native core banking solutions

- Member of the corporate infrastructure team aimed at developing the tools to enable deployment, testing and integration to increase developers speed.
- Contributed writing and improving the internal **Python CLI** to release and cloud deployment.
- Wrote and managed **jenkins** deployment **cron jobs**.
- Wrote a **Python** service to scrape **Phabricator** and sync its tickets into **JIRA**.

SEPTEMBER 2014 – SEPTEMBER 2018

MRes + PhD in medical image analysis | [UCL](#)| London, UK

Research student, CDT program (funded MRes + PhD) in medical imaging and bioengineering

- Implemented **ML models** and automated statistical analysis pipelines to quantify the negative effects of steroids administration in preterm birth, as part of a multi-disciplinary international research team.
- Developed a novel numerical analysis method to integrate **ODE** in diffeomorphic image registration.
- Published **7 peer reviewed papers** also on Neuroimage and Nature scientific report, about diffeomorphic image registration and Machine Learning to automate high resolution magnetic resonance images segmentation.
- Open sourced 12 Python libraries and one Micro MRI template dataset of 12 subjects manually segmented.

MARCH 2013 – JUNE 2014 (LEFT AFTER WINNING A PHD STUDENTSHIP AT UCL)

Industrial Simulation Modeller | [SimTec](#) | Turin, Italy

Discrete Event simulation for the Automotive Industry

- Developed material flow simulation models with **PlantSimulation** and **SimTalk** to estimate efficiency, remove bottlenecks and dimension buffers.
- Supported industrial plant layout design for a range of clients in Italy and Germany.
- Developed in-house shortest path algorithms for the internal and external logistics of assembly parts, to reduce

lags in JIT manufacturing.
• Presented my results at the first annual Tecnomatix Plant Simulation User Conference in Stuttgart.

JUNE 2011 – OCTOBER 2011
Developer | [TcWeb](#) | Turin, Italy
Web development and technology consulting

• Junior developer, Java J2EE and Struts 2 for developing the website of Regione Piemonte
• Document existing code with UML diagrams
• Prototyped and implemented a generalized version of the Hungarian Algorithm to digitalise newspaper pages, reducing 2 months of manual work in less than 1 minutes of computations.

Selected Publications

• S Ferraris [Bourbaki vs Pragmatism: A methodological comparison through the multi-armed bandits problem](#) - preprint 2023
• Ferraris S, Lorenzi M, Daga P, Modat M, Vercauteren T. [Accurate small deformation exponential approximant to integrate large velocity fields: Application to image registration](#) IEEE CVPR 2020.
• van der Merwe J, van der Veeken L, Ferraris S, et al. [Early neuropathological and neurobehavioral consequences of preterm birth in a rabbit model](#) Nature scientific reports, May 2019.
• Ferraris S, van der Merwe J, et al. [A magnetic resonance multi-atlas for the neonatal rabbit brain](#) Neuroimage, 2018.
• Ferraris S, Shakir ID, Van Der Merwe J, et al. [Bruker2nifti: Magnetic resonance images converter from Bruker ParaVision to NIfTI format](#) Journal of Open Source Software 2017.

Skills

Data science • Research • data pipeline automation • prototypes and POC • supervised and unsupervised ML • mathematical modeling • performance analysis • bottleneck analysis • design and packaging python libraries • production-grade quality Python (scikit-learn, pandas, numpy, streamlit, DeckGl, KeplerGl) • Git (GitHub, GitLab) • Docker • CI/CD and automation • unit testing • contract testing • Airflow and Astronomer • AWS • Fast API • Databricks.

Education

2014 - 2018 - **MREs + PhD, Centre for Doctoral Training (EPSRC), Medical Imaging | UCL | London, UK**
2010 - 2013 - **Master of Science, Mathematics | Università degli Studi di Torino | Turin, Italy**
2006 - 2010 - **Bachelor's of Science, Mathematics | Università degli Studi di Torino | Turin, Italy**

Activities

Volunteering: Maths tutor for Action Tutoring, Scanner and Marshall for ParkRun • Classical guitar player • Cross country running