Sebastiano Ferraris

Data Scientist, PhD

London

sebastiano.ferraris@gmail.com

github.io/GeoDsBlog/

github.com/SebastianoF

linkedin.com/in/ibis-redibis/

Google Scholar

Research Gate

Skills

Python 9+ years

Data science 9+ years

Algorithms 9+ years

Artificial intelligence 4+ years

Geospatial data science 4+ years

Medical image analysis 4 years

Discrete events simulation 1 year

Dynamic pricing 1 year

Summary

5+ years experience in developing 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 around productionisation, deployment and algorithms continuous validation. Scientific author published in international journals.

Experience

Data scientist | **General System** — **June 2020 - June 2024**

Startup in stealth model until 2022. High performant real-time analytic platform for high volume (100+Bn) spatiotemporal data *Startup declared insolvency in June 2024.*

- Wrote and tested 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. Create ETL and 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**, **Open API**, **unittesting**, **integration testing**.
- Developed and open sourced a python library to provide users tooling and examples for the Data Flow Index.
- Contributed to the company blog aimed at building a community around the topics of spatiotemporal data science.
- Supported customer success and marketing with visualizations and presentation materials.

Algorithm engineer | **Pace** — **Sept 2019 - June 2020**

Predictive analytics and dynamic pricing for the hospitality industry integrated in the PMS. *Left due to COVID-19 disruption in the hospitality industry*

- Participated in developing an **agent based simulation** aimed at validating the core prediction algorithm.
- Maintained Python and SQLAlchemy production code with the BE team.
- Migrated production codebase to **Dask** to improve scalability.

Back end developer | **Thought Machine** — **Oct 2018 - June 2019**

Cloud native core banking. *Left to pursue a career in algorithms development and data science more aligned with my studies.*

- 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 for releasing and cloud deployment.
- Wrote and managed jenkins deployment cron jobs.
- Wrote a Python service to scrape Phabricator and sync its tickets into JIRA.

Education

2015 - 2018

PhD, Centre for Doctoral Training (EPSRC), Medical Imaging

University College London

MRI ▪ Pre-clinical studies ▪ Numerical methods for Image registration ▪ 8 Papers published ▪ 12 repositories open sourced

2014 - 2015

Master of Research (MRes), Medical Imaging

University College London

Numerical methods for image registration ▪ Digital Image Processing ▪ Optics in Medicine

2010 - 2013

Master of Science (MSci), Mathematics

Università degli studi di Torino

Geometry ▪ Error correcting code theory ▪ Computational modelling.

2006 - 2010

Bachelor's of Science (BSc), Mathematics

Università degli studi di Torino

Volunteering

- Maths Tutor, Action Tutoring
- Scanner and Marshall, Parkrun

MRes + PhD in medical image analysis | UCL — **Sept 2014 - Sept 2018**

Research Student

- 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 for automated MRI segmentation.
- Reproducible research advocate: open sourced 12 Python libraries (Sec 7.2.2 of my PhD Thesis) and one micro MRI dataset.

Industrial simulation modeller | SimTec — **March 2013 - June 2014**

Automotive industry, discrete events simulation

- Developed material flow simulation models with PlantSimulation (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.

Developer | TcWeb — **June 2011 - Oct 2011**

Web development and technology consulting

- Junior developer, Java J2EE and Struts 2 for developing the website of Regione Piemonte.
- Document existing codebase 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

- Ferraris S, et al. "A magnetic resonance multi-atlas for the neonatal rabbit brain". Neuroimage. *Neuroimage* 2018 Oct doi: 10.1016/j.neuroimage.2018.06.029.
- Ferraris S, et al. "Accurate small deformation exponential approximant to integrate large velocity fields: Application to image registration". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Lipsum, June 12-17, 2020.
- Ferraris S, et al. "Bruker2nifti: Magnetic resonance images converter from bruker ParaVision to NIfTI format". In: *Journal of Open Source Software*, 2017.
- Ferraris S "Image computing tools for the investigation of the neurological effects of preterm birth and corticosteroid administration" *PhD thesis, University College London*, 2019.

1st July 2024, Online version link.