Sebastiano Ferraris

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

Linkedin • Github • Blog • Google Scholar • Research Gate

Data Scientist and Researcher 5+ years of 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 and algorithms continuous validation. Scientific author published in international journals.

Skills

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

Experience

June 2024 – Present

Senior Data Scientist | Kpler | London, UK

Leading provider of technology-led data, analytics, and market insights focused on the energy and shipping markets.

June 2020 – June 2024

Data Scientist | General System | London, UK

Startup in stealth mode 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 with 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 the algorithms described above to detect: dark vessels, crowds gathering, consumers' patterns and cross

visitations. Set 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 and integration testing.

- Supported customer success and marketing with visualizations and presentation materials.
- Open sourced a python library and a series of examples for analysts to interact with the General System's platform.
- . Contributed to the company blog and presented prototypes and findings to stakeholders and potential clients.

September 2019 – June 2020

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 Back end developer | <u>Thought Machine</u> | 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

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.

Selected Publications

- · Bourbaki vs Pragmatism: A methodological comparison through the multi-armed bandits problem preprint 2023
- Accurate small deformation exponential approximant to integrate large velocity fields CVPR 2020.
- Early neuropathological and neurobehavioral consequences of preterm birth in a rabbit model Nature scientific reports, 2019.
- <u>A magnetic resonance multi-atlas for the neonatal rabbit brain</u> Neuroimage, 2018.
- Bruker2nifti: Magnetic resonance images converter from Bruker ParaVision to NIfTI format JOSS 2017.

Education

- 2015 2018 PhD, Centre for Doctoral Training (EPSRC), Medical Imaging | UCL London, UK
- 2014 2015 MREs, Centre for Doctoral Training (EPSRC), Medical Imaging | UCL London, UK
- 2010 2013 MSCi, 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
- Runner