STEPHEN RICHER

+44(7513) 065 917 ◇ London, U.K.

stephen.richer@proton.me ◇ linkedin.com/in/stephenricher ◇ github.com/StephenRicher

OBJECTIVE

Python programmer and data scientist with an extensive quantitative research background in biology and mathematics. Since moving to an industry data science position, within the U.K. National Health Service (NHS), I have worked to develop best practise approaches for building, packaging and deploying machine learning workflows and analytical tools within the NHS. My work has focussed on identifying and addressing patterns of health inequality within the healthcare service. Having thoroughly enjoyed coaching quantitative research skills throughout my PhD, I am now seeking an opportunity to build upon and share my expertise with others.

SKILLS

Programming Languages	Python (NumPy, Pandas, Scikit-learn), SQL, Git, Bash, C++, Java
Data Science	Supervised Machine Learning (Decision Trees, Regression), Tableu
Cloud Platforms	Microsoft Azure
Soft Skills	Communication, Critical Thinking, Problem Solving, Adaptability

EXPERIENCE

Data Scientist Intern   Jun 2022 - Dec 2022
NHS England   *London, U.K*

- Built and deployed an AutoML pipeline, implementing CatBoost and Logistic Regression, for predicting healthcare appointment non-attendance at regional NHS trusts. Available at: github.com/nhsx/dna-risk-predict
- Developed ETL data engineering pipelines for processing and aggregating healthcare and demographic data from multiple public source.
- Collaborated with NHS stakeholders and analysts to design and implement reproducible analytical workflows for studying healthcare inequalities.
- Utilised a variety of statistical approaches and hypothesis tests, including permutation testing and partial correlation, to robustly analyse complex and biased real-world healthcare data with numerous confounding variables.

Research Software Engineering Coach   Apr 2019 - Jun 2022
University of Bath, Doctoral College   *Bath, U.K*

- Developed and delivered materials to train doctoral students in research software engineering best practises.
- Curriculum: Python programming and style (PEP8), version control, testing and continuous integration. Available at: github.com/Research-Software-Skills-Bath

Quantitative Research Scientist, Biology & Mathematics   Oct 2018 - Jun 2022
University of Bath   *Bath, U.K*

- Conducted independent research and developed novel computational tools to understand how the structure of DNA impacts cellular function.
- Developed HiCFlow, a user-friendly analytical workflow for conducting robust and reproducible bioinformatics analysis. Available at: github.com/StephenRicher/HiCFlow
- Supervised final-year undergraduate students completing projects in computational biology.
- Presented research at national and international conferences.
- Awarded best final-year PhD presentation at the Bath Departmental Research Day.

Technical Support (IT Services)   Oct 2017 - Sep 2018
University of Manchester, Hornet   *Manchester, U.K*

- Provided IT support, education and security guidance for residential students.

Study Coordinator Oct 2016 - Jul 2017

MAC Clinical Research *Leeds, U.K*

- Oversaw a phase 3 clinical trial investigating a novel treatment for Alzheimers disease.

Clinical Trials Assistant Oct 2015 - Oct 2016

MAC Clinical Research *Leeds, U.K*

EDUCATION

PhD, Biology & Mathematics, University of Bath 2018 - 2022

Mathematical and bioinformatics-based tools to explore the impact of gene editing on the geometric principles governing the 3D structure of the genome.

MSc, Bioinformatics & Systems Biology (Distinction), University of Manchester 2017 - 2018

BSc, Biology (First), University of Bath 2011 - 2015

PROJECTS

Real World Data Validation with *validatum* Lightweight, user-friendly Python module to automatically detect and flag suspicious data inconsistencies and common data entry errors within real world data.
Available at: pypi.org/project/validatum

Secure Data Encryption with *datasafe.* Python based command-line utility for encryption of text files and encryption of Pandas DataFrames while preserving datatype.
Available at: pypi.org/project/datasafe

Data Science Best Practises - The Titanic Dataset. Wrote a popular Kaggle notebook, describing best practise approaches for performing supervised tabular classification using Scikit-learn.
Available at: Kaggle - Titanic Data Science