

SUMMARY

Data Scientist particularly interested in **imaging**, **geospatial data**, and, **text-mining**, with two years experience delivering commercial data science projects. Dedicated to applying machine learning to solve problems with a positive social impact, and fascinated to apply the latest deep learning methods, and explore how new technologies can be used to benefit society. Hold a **PhD in Bioinformatics**, and a **MSc in Computer Science**, with over seven years programming experience.

PROFESSIONAL EXPERIENCE

Data Science for Social Good (DSSG)

Data Scientist

Jun 2017 – present

Completed summer fellowship, and selected to help establish DSSG Europe group.

– Developed an open-source risk tool combining vessel tracking data with satellite imagery. Implemented a machine learning model to classify fishing vessels, and then score vessels according to likely illegal fishing behaviours, based on multiple risk indicators. Project code is mostly available here: https://github.com/DSSG2017/wef_oceans

– Lead a team in creating a data visualisation platform working with the Netherlands Rijkswaterstaat transportation ministry. Deployed a PostgreSQL database on Azure, and performed data ETL, cleaning, and visualisation of traffic flows, speeds, and accidents across 12 major border crossings.

ASI Data Science / London Fire Brigade

Data Science Consultant

Jan 2017 – Apr 2017

Advanced Skills Initiative - intensive post-doctoral training.

– Completed intensive 8 week training in machine learning, databases, and statistics, as well as business skills including: communication, negotiation, project management, and commercial awareness.
– Worked in house with the London Fire Brigade to implement topic modelling methods to classify corpus of 37,000 fire incident reports. In the process revealed previously unknown incident types, and visualised results. Project code: https://github.com/williamgrimes/london_fire_brigade

LMCB / A*STAR Bioinformatics Institute

PhD Candidate

Sep 2013 – Nov 2016

Joint programme at the Laboratory for Molecular Cell Biology, London, and A*STAR, Singapore.

– Applied machine learning techniques to identify cell phenotypes with a detection accuracy of 82%
– Published articles in *Nature Scientific Reports* and *Journal of Thrombosis and Haemostasis*.

National Institute of Informatics / 国立情報学研究所

International Internship Programme

Jun 2013 – Sep 2013

– Developed a Java plugin to aid geneticists annotating phenotypes in μ CT images of mouse embryos.

EDUCATION

University College London

PhD of Bioinformatics

Sep 2013 – Nov 2016

- *Image processing and analysis methods in quantitative endothelial cell biology*

University College London

MSc of Computer Science

Sep 2012 – Sep 2013

- Awarded Aadrvarik Scholarship 2012

Durham University

Natural Sciences – Physics and Geophysics

Sep 2009 – Sep 2012

SKILLS / KEYWORDS

Programming: Python, R, Java, Matlab, SQL, Solidity

Machine Learning: regression, decision trees, ensembles, SVM, neural networks, deep learning

Libraries: NumPy, pandas, SciPy, scikit-learn, scikit-image, OpenCV, Keras, theano, tensorflow

Visualisation: matplotlib, bokeh, plotly, ggplot, Tableau

Scripting: Bash and utils (sed, awk, grep), Python, Ruby

Databases: MySQL, PostgreSQL, SQLite

Misc: Amazon Web Services, Azure, Docker, Git, L^AT_EX, UNIX