

## SUMMARY

**Data Scientist**, with two years experience delivering commercial data science projects working with the *World Economic Forum*, *Netherlands' Transportation Ministry*, and the *London Fire Brigade*. Hold a **PhD in Bioinformatics**, and a **MSc in Computer Science**, with over eight years programming experience. Dedicated to applying machine learning to solve problems with social impact, and particularly interested in *imaging*, *geospatial data*, and *natural language processing*.

## PROFESSIONAL EXPERIENCE

### Numerico Technologies

*Data Scientist and Engineer*

**Feb 2018 – present**

Data Scientist and software-engineer.

→ Developed and deployed a live system for the Netherlands Rijkswaterstaat transportation ministry using neural networks to predict the likelihood of road traffic accidents along Dutch highways.

### Data Science for Social Good (DSSG)

*Data Scientist and Mentor*

**Jun 2017 – Jan 2018**

DSSG Europe 2017 member and DSSG 2018 technical mentor.

→ Developed an open-source fishing risk tool combining vessel tracking data with satellite imagery, using a random forest to score vessels according to likelihood of illegal fishing behaviours. Project code is available here: [https://github.com/DSSG2017/wef\\_oceans](https://github.com/DSSG2017/wef_oceans)

### ASI Data Science / London Fire Brigade

*Data Science Consultant*

**Jan 2017 – Apr 2017**

Fellowship programme in commercial data science for researchers with strong analytic background.

→ In house consultant with the London Fire Brigade, implementing topic modelling methods to classify corpus of 37,000 fire incident reports. Revealed previously unknown incident types, and visualised results. Project code: [https://github.com/williamgrimes/london\\_fire\\_brigade](https://github.com/williamgrimes/london_fire_brigade)

### Laboratory for Molecular Cell Biology (LMCB)

*PhD Student*

**Sep 2013 – Nov 2016**

Joint scholarship at LMCB and A\*STAR Bioinformatics Institute, Singapore.

→ Applied machine learning techniques and segmented endothelial cell phenotypes in high-throughput microscopy assays. Analysed over 5TB of data with a detection accuracy of 82%.

### National Institute of Informatics (国立情報学研究所)

*International Internship Programme*

**Jun 2013 – Sep 2016**

→ Developed Java-based annotation software which assisted in classifying developmental phenotypes in  $\mu$ CT images of mutant mice

## EDUCATION

### University College London

*PhD Bioinformatics*

**Sep 2013 – Nov 2016**

→ Thesis: *Image processing and analysis methods in quantitative endothelial cell biology*

### University College London

*MSc Computer Science*

**Sep 2012 – Sep 2013**

→ Awarded Aardvark Scholarship 2012

### Durham University

*BSc Natural Sciences – Physics and Geophysics*

**Sep 2009 – Sep 2012**

## SKILLS / KEYWORDS

**Programming:** Python, R, Java, Matlab, SQL, Solidity, Bash and utils (sed, awk, grep)

**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

**Databases:** MySQL, PostgreSQL, MS SQL Server

**Front-end:** Javascript, HTML, CSS, Semantic UI, Jekyll

**Misc:** Amazon Web Services, Azure, Docker, Git, Jupyter Notebooks, L<sup>A</sup>T<sub>E</sub>X, UNIX