Laurens Geffert, PhD

Experience

Jan 2017 – present
London, UK

Data Scientist at Sainsbury's Data and Analytics Centre of Excellence (DACE)
Leading UK retailer putting data and analytics at the centre of decision making on all levels

- Created models for procurement and supply chain optimisation using graph theory, linear optimisation, and machine learning methods in R, Python, and Spark
- Identified value opportunities worth tens of millions of pounds in cost savings, of which £1.3 million per year were immediately realised
- Successfully communicated complex findings to a range of business stakeholders
- Planned and supervised the work of two junior analysts on multiple projects

Jan 2016 – Jan 2017
(1 year)
London, UK

Data Scientist at VisualDNA / Coremetrix
Financial technology company using psychometric scoring in the credit & risk sector

- Built predictive credit risk models using psychometric data and bureau data
- Scaled up risk models using event tracking from millions of customers with Hive and R
- Designed and implemented a new data pipeline and migrated legacy data to AWS
- Developed new methods to predict motor insurance risk using psychometric tests, providing a 100% uplift in predictive performance on existing models
- Delivered proof of concept work for of a new client that led to a contract worth over £300k to the business

Oct 2015 – Nov 2015
(2 months)
London, UK

Data Science Intern at Science to Data Science (S2DS) / OpenSensors
Conversion course and data science boot-camp with an IoT start-up

- Developed a full-stack solution for sensors deployed as part of a citizen science project
- Provided analytical insights to airport, politicians, and local community

Sep 2011 – Oct 2015
(4 years)
Cambridge, UK

Researcher at UNEP-WCMC
UN branch and world-leading think tank for biodiversity policy and decision making

- Wrote scientific reports for UN bodies, EU commission, and CITES secretariat
- Trained other staff members in python, GIS, and spatial analysis
- Automated data input for legacy in-house databases using Python, replacing manual labour and speeding up the process from a minute per entry to under a second

Mar 2009 – Aug 2011
(2 years 6 months)
Bonn, Germany

Scientific Research Assistant at the Nees-Institute, University of Bonn
Academic research institute gathering, mapping, and analysing biodiversity data

- Modelled the distribution of West African plant species using machine learning
- Automated complex workflows using ArcGIS, R, and Python

Education

Oct 2012 – Dec 2015
(3 years 3 months)
Cambridge, UK

PhD at the University of Cambridge
Thesis: Predicting the Global Distribution of Commercially Important Marine Species

- Created spatial predictions from noisy and biased data using machine learning
- Developed an approach to correct for bias in the model training data
- Improved the predictive performance of models by 22%
- Scaled models to tune and run for 1000s of species in parallel in an HPC environment

Oct 2004 – Jun 2011
(6 years)
Bonn, Germany

Diplom (~MSc equivalent) Student of Biology at the University of Bonn
Thesis: Spatial Patterns of Moss Species Diversity on a Global Scale

- Focused on macroecology, numerical ecology, and biodiversity informatics
- Applied data mining (OCR, web scraping, online API queries) to collate new datasets
- Generated insight from data using multivariate statistics (linear regressions, ordination)
- Published results in three papers and presented at four international conferences

Laurens Geffert, PhD

Programming Languages

R (7 years experience)
SQL (2 years experience)
Python (2 years experience)
JavaScript (1 year experience)

Natural Languages

German (native)
English (fluent)
Spanish (working)

Technical Skills

Some of the methods and programmes I have used in the past

Statistical Modelling
- Variable Selection (e.g. Stepwise, L1 regularisation)
- Model Evaluation (e.g. RMSE, AUC)
- Cross-validation

Mathematical modelling
- Linear programming
- Optimisation

Data Management
- SQL
- HDFS
- Hive
- Spark

Machine learning
- Supervised
  - Linear and Logistic Regression
  - Random Forest
  - Boosted trees
- Unsupervised
  - k-Means
  - NMDS
  - t-SNE

Geospatial Data
- Raster and Vertex Data
- Spatial Analysis
- Spatial Visualisation

Achievements

First prize winner at ECMWF hackathon 'Beyond weather: explore creative uses of open data' (2017)
Built a predictive model for assessing risks of bus line breakdowns and accidents

Third prize winner of the GBIF Ebbe Nielsen Challenge (International hackathon by GBIF, 2016)
Wrote an R-package that allows users to correct for sampling bias in large biological datasets

Captain of Darwin College Boat Club (2014 – 2015)
Chaired a committee of ten, organised daily training sessions and weekly competitions for ~100 members, won multiple competitions and increased the club's operational budget by 75%

Voted 'Most valuable player' at the World-class University Rowing Festival, Daegu, South Korea (2015)

Contributor to 'rgbif', an R-package with over 40,000 downloads