Jori van Lier – Data Scientist

Location:	Leiden, The Netherlands
Gender:	Male
Nationality:	Dutch
Email:	info@jorivanlier.nl
Tel:	+31 (0)6 5207 8873

I am a freelance data scientist with a never-ending curiosity. My speciality is machine learning: I am familiar with a wide range of techniques to explore datasets and to build predictive models. Next to that I have a strong foundation in engineering and distributed real-time systems. I'm more than happy to put in the engineering effort to deploy an analysis or model to production at scale. I am comfortable in a senior role, defining tasks, priorities and coaching junior teammates, but still love to dive into datasets myself as well.

Work experience

Oct 2017 – Onwards: Freelance Data Scientist

Helping my clients extract value from data as an independent data scientist.

Oct 2016 – Sep 2017: KPMG Advisory – Manager Big Data & Analytics

Functioning as a senior data scientist, I was responsible for analytics at BlueSense, a start-up that was incubating within KPMG. BlueSense is a platform that monitors the behaviour of people within buildings, in real-time, using various sensors. It includes state of the art analytics to understand crowd density, crowd movement patterns, dwell times and occupancy levels, both descriptive and predictive. This system was deployed at various clients in retail, public transport, facility management and a football stadium.

Dec 2013 – Sep 2016: KPMG Advisory – Senior Consultant Big Data & Analytics

I was lucky enough to join a team of big data experts with a mission to disrupt industries with novel data-driven approaches. We were one of the pioneers in the big data space, bringing machine learning, simulations and statistics to regular businesses. I did a mix of client projects and product development for a real-time Location Analytics platform (which would later evolve into BlueSense).

2012 – 2013: KPMG Advisory – Senior Consultant Risk Consulting

Business intelligence and data quality projects.

2010 – 2012: KPMG Advisory – Consultant Risk Consulting

Business intelligence and data quality projects.

2007 – 2010: Webtechniek – Developer (part time)

Developing various PHP/MySQL web applications.

Education

MSc Computer Science, 2008 – 2010

Leiden University. Thesis: "*Trading strategy optimization using evolutionary algorithms*". Graduated cum laude.

BSc Computer Science, 2004 – 2008

Leiden University.

Various courses

Management Development, KPMG, 2017.
Calculus Two, Ohio State University via Coursera, 2017.
3-day advanced statistics course, Nikhef (the Dutch institute for subatomic physics), 2016.
Calculus One, Ohio State University via Coursera, 2015.
Machine Learning, Stanford University via Coursera, 2014.
Statistics, Duke University via Coursera, 2014.

Skills and Competences

Languages: Dutch (native), English (fluent).

Programming languages: Highly proficient with Python and Java. Average knowledge of JavaScript. Familiar with the basics of Scala. Deeply familiar with the following Python packages: numpy, pandas, scikit-learn, scipy, matplotlib, pycm3, keras, flask.

Big Data platforms: Hadoop (Hortonworks), Hive, YARN, HDFS, Spark, Storm, Kafka.

Web: HTML, CSS, Bootstrap, JQuery, a little D3.js.

Databases: SQL in various dialects, mostly MySQL, PostgreSQL and Hive. NoSQL: MongoDB.
Dev tools: Git, Maven, Archiva, Jenkins, Docker, Vagrant.

Linux: CentOS, Ubuntu, Bash scripting, Amazon Web Services.

Machine learning: Familiar with too many methods to describe here, but most of my experience involves linear models, logistic regression, (stochastic) gradient descent, decision trees, random forests, boosted trees, k-means clustering, (convolutional) neural nets, backprop, PCA, k-fold cross-validation, learning curves, ROC curves, bias-variance dilemma, ensembles.

Statistics: Familiar with both frequentist and Bayesian approaches. I take a simulation based approach to frequentist hypothesis testing and uncertainty estimation (bootstrap). Good understanding of probability theory, likelihood methods, confidence intervals, chi^2 tests, EM, Bayes rule, Bayesian credible intervals, Bayes factor. Fascinated by Bayesian methods using MCMC.