

Jeroen Janssens

Personal information

Website: jeroenjanssens.com
GitHub: github.com/jeroenjanssens
LinkedIn: linkedin.com/in/jeroenjanssens
Email: jeroen@jeroenjanssens.com
Location: Rotterdam, the Netherlands

Summary

As an independent consultant and trainer, I help organizations making sense of their data. I'm also an assistant professor of data science at Tilburg University. Previously, I was a data scientist at Elsevier in Amsterdam and startups YPlan and Visual Revenue in New York City. I have an MSc in artificial intelligence from Maastricht University and a PhD in machine learning from Tilburg University. Recently, I wrote a book titled *Data Science at the Command Line*, which was published by O'Reilly.

Work experience

Independent consultant and trainer

Rotterdam
October 2014 – present

As a consultant, I help organizations gain more insight, make better decisions, and build smarter products. I either implement a stand-alone project, temporarily support an existing team, or set up a data science practice from scratch. In addition, as a trainer, I provide in-house workshops for data scientist and engineers about topics such as Machine Learning, R, Python, and Linux. Clients include: KPN, Snow, Social Point, Jheronimus Academy of Data Science, and Container Solutions.

Assistant professor of data science

Tilburg University
November 2015 – present

For one day a week, I work at Tilburg University for the master Data Science: Business and Governance. I teach the course *Statistical Programming with R* and supervise several students with their master theses.

Lead data scientist

Elsevier, Amsterdam
December 2014 – July 2015

Elsevier is a provider of scientific, technical, and medical information products.

- Set up a new project to build a platform for alerts and recommendations that will eventually be employed by all products within the RELX Group
- Analyzed and visualized large text corpora
- Collaborated with the University of Amsterdam to implement a search query recommender

Author

O'Reilly Media, New York City
January 2014 – November 2014

Authored a book titled *Data Science at the Command Line*, which was published by O'Reilly in October 2014. From the cover: "This hands-on guide demonstrates how the flexibility of the command line can help you become a more efficient and productive data scientist. You'll learn how to combine small, yet powerful, command-line tools to quickly obtain, scrub, explore, and model your data." So far, the book has been translated to Chinese and Japanese; currently we're working on both a Russian and a Spanish translation.

Senior data scientist

YPlan, New York City
October 2013 – October 2014

YPlan is an app that lets you browse through a curated list of events (e.g., concerts, comedy shows, and wine tastings) and book them in two taps.

- Led the realization of a content-based recommendation system
- Implemented a backtesting framework to evaluate the performance of the recommendation system on historical data
- Performed ad-hoc data analysis for the sales and marketing teams

Data scientist

Visual Revenue and Outbrain, New York City
January 2012 – October 2013

Visual Revenue offers tools for news editors to optimize their front pages and social media channels. Customers include: Forbes, NY Daily News, Le Monde, and De Telegraaf. In March 2013, Visual Revenue was acquired by Outbrain.

- Analyzed and visualized real-time streaming data
- Developed a Bayesian A/B headline testing tool
- Automated archival of streaming data to AWS S3
- Implemented an evolutionary algorithm that optimizes the time to post on social media

Education

Tilburg University

Ph.D. in Machine Learning
January 2008 – January 2013

The focus of my thesis, titled *Outlier Selection and One-Class Classification*, is on automatically detecting anomalies in large data sets using statistics and machine learning algorithms. Supervisors: prof.dr. Eric Postma and prof.dr. Jaap van den Herik. I visited the Computational and Biological Learning Lab at the University of Cambridge for four months to work with prof. Zoubin Ghahramani on mixed-membership models.

Machine Learning Summer School

Organised by the University of Cambridge, Microsoft Research, Pascal
29 August 2009 – 10 September 2009

Received two weeks of training on topics such as Bayesian statistics, stochastic optimisation, probabilistic graphical models, computer vision, and information retrieval.

Maastricht University

Master of Science in Artificial Intelligence (*cum laude*)
September 2006 – January 2008

Included courses such as Information Retrieval, Multi-Agent Systems, and Games & AI. Thesis title: *Collaborative Image Ranking: Bridging the Semantic Gap using Human Computation*.

University College Maastricht (UCM)

Bachelor of Science (*cum laude*)
September 2003 – June 2006

UCM is a liberal arts college, which allowed me to follow a variety of courses, including Data Mining, Linear Programming, Knowledge Management, and Bioinformatics, but also courses such as Philosophy and Psychology.

University of Adelaide, Australia

Semester abroad as part of UCM
June 2005 – January 2006

Studied one semester of the third year of the Computer Science program at the University of Adelaide. Courses included Distributed Systems, Software Engineering, and Programming in C and C++.

Selected publications

- J.H.M. Janssens. *Data Science at the Command Line: Facing the Future with Time-Tested Tools*. O'Reilly Media, Sebastopol, CA, 2015
- J.H.M. Janssens, F. Huszar, E.O. Postma, and H.J. van den Herik. Stochastic Outlier Selection. Technical Report TiCC TR 2012-001, Tilburg University, Tilburg, The Netherlands, 2012
- J.H.M. Janssens, H. Hiemstra, and E.O. Postma. Creating artificial vessel trajectories with Presto. In *Proceedings of the 22nd Benelux Conference on Artificial Intelligence*, Luxembourg, Oct. 2010
- C.M.E. Willems, W.R. van Hage, G.K.D. de Vries, J.H.M. Janssens, and V. Malaisé. An integrated approach for visual analysis of a multi-source moving objects knowledge base. *International Journal of Geographical Information Science*, 24(10):1543–1558, 2010

Skills and competences

I have experience and affinity with:

- Analyzing data using R and the PyData stack
- Tinkering with machine learning algorithms
- Organizing workshops and meetups
- Writing and presenting for technical audiences
- Managing data in MySQL, Mongo, Redis, and S3
- Programming in Python, JavaScript, and Bash
- Visualizing data with ggplot and D3
- Performing large, distributed experiments