

Jeroen Janssens, Ph.D.

About.

Jeroen Janssens is an independent data science consultant and instructor. He enjoys visualizing data, implementing machine learning models, and building solutions using Python, R, JavaScript, and Bash. He's passionate about helping others to do such things.

Jeroen manages Data Science Workshops, a training and coaching firm that organizes open enrollment workshops, in-company courses, inspiration sessions, hackathons, and meetups. Clients include Amazon, DPD, eHealth Africa, KPN, Schiphol Airport, and The New York Times.

Previously, he was an assistant professor at Jheronimus Academy of Data Science and a data scientist at Elsevier in Amsterdam and various startups in New York City. He is the author of *Data Science at the Command Line*, published by O'Reilly Media. Jeroen holds a PhD in machine learning from Tilburg University and an MSc in artificial intelligence from Maastricht University.

He lives with his wife and two kids in Rotterdam, the Netherlands.

Contact.

Jeroen is available to provide consulting and training in the areas of data science, data engineering, and machine learning. He's also available to speak at private and public events. If you would like to know more about his services, fees, and availability, then please [email Jeroen](#). He can be reached at:

- Email: jeroen@jeroenjanssens.com
- Personal website: jeroenjanssens.com
- Company website: datascienceworkshops.com
- LinkedIn: linkedin.com/in/jeroenjanssens
- GitHub: github.com/jeroenjanssens
- Twitter: twitter.com/jeroenhjanssens

Experience.

Founder and Instructor

Data Science Workshops B.V.
Jan 2017 – Present
Rotterdam

Data Science Workshops is a leading training and coaching firm, providing hands-on workshops, inspiring masterclasses, and complete in-company courses in subjects such as data visualization, programming, and machine learning. Clients include Amazon, DPD, eHealth Africa, Schiphol Airport, and The New York Times.

Independent Data Science Consultant

Jeroen Janssens
Jul 2015 – Present
Rotterdam

Implemented data science analyses and solutions for a variety of clients, such as KPN, Teradata, Telenor, and several startups. Some projects were entirely remotely, while others were on-site.

- Analyzing and visualizing large amounts of high-dimensional data
 - Building feature extraction pipelines for natural language processing
 - Converting a machine learning algorithm from MATLAB to Python
 - Implementing RESTful APIs for online prediction
 - Developing a dashboard in R Shiny
-

Author

O'Reilly Media
Jan 2014 – Present
New York City / Rotterdam

Authored *Data Science at the Command Line*, a hands-on guide that demonstrates how the flexibility of the Unix command line can help you become a more efficient and productive data scientist. Published by O'Reilly in October 2014. On datascienceatthecommandline.com, the book can be read for free. I'm currently working on the second edition, which is scheduled to be published in Q3 2021.

Organizer

Data Science NL Meetup

Mar 2019 – Present

Amsterdam

Data Science NL is a community with over 3,000 members dedicated to sharing knowledge about the exciting multidisciplinary field of data science. Hosts include Heineken, Uber, Picnic, and TomTom.

Assistant Professor of Data Science

Jheronimus Academy of Data Science / Tilburg University

Nov 2015 – Sep 2017

Den Bosch / Tilburg

Supervised several master students with their thesis. Taught the course *Introduction to Machine Learning* as part of the Data Science pre master at Jheronimus Academy of Data Science. Prior to that I taught the course *Statistical Programming with R* as part of Tilburg University's Data Science: Business and Governance Master program.

Lead Data Scientist

Elsevier

Dec 2014 – Jul 2015

Amsterdam

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 would eventually be employed by all products within the RELX Group
 - Managed a team of engineers to implement a prototype of said platform
 - Analyzed and visualized large text corpora using Apache Spark
 - Collaborated with the University of Amsterdam to implement a search query recommender
-

Senior Data Scientist

YPlan

Oct 2013 – Oct 2014

New York City

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 / Outbrain

Jan 2012 – Oct 2013

New York City

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. I was the first data scientist.

- Interviewed, hired, and on-boarded new data scientists
 - Analyzed and visualized real-time streaming data
 - Developed a Bayesian A/B headline testing tool, a primary feature of the platform
 - Automated the long-term storage of streaming data to AWS S3 using Python
 - Implemented an evolutionary algorithm that optimizes what to post on social media
-

Education.

Tilburg University

Ph.D. in Machine Learning

Jan 2008 – Jan 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.dr. Zoubin Ghahramani on mixed-membership models.

Machine Learning Summer School

Organised by the University of Cambridge, Microsoft Research, Pascal
Aug 2009 – Sep 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)
Sep 2006 – Jan 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)
Sep 2003 – Jun 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
Jun 2005 – Jan 2006

Studied a 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 and Talks.

- *Visualizing High-Dimensional Data with Python*. O'Reilly Live Training. June 15, 2020.
- J.H.M. Janssens. *Plotnine: Grammar of Graphics for Python*. A translation of the visualisation chapters from "R for Data Science" to Python using Plotnine and Pandas. December 11, 2019
- *Scalable Anomaly Detection With Spark and SOS*. Strata Data Conference. New York, NY. September 26, 2019
- J.H.M. Janssens. *Data Science at the Command Line: Facing the Future with Time-Tested Tools*. O'Reilly Media, Sebastopol, CA, 2015
- *Building a Data Science Toolbox*. Data Science London Meetup. London. April 10, 2014
- J.H.M. Janssens. *Outlier Selection and One-Class Classification*. PhD thesis. Supervised by Eric Postma and Jaap van den Herik. Tilburg University, June 11, 2013
- 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 Competencies.

I have experience and affinity with:

- Collaborating in remote teams
- Visualizing data with ggplot and D3
- Organizing workshops and meetups
- Using tools such as Git, Docker, and Make
- Performing large, distributed experiments
- Developing packages in both R and Python
- Managing data in SQL, Mongo, Redis, and S3
- Writing and presenting for technical audiences
- Programming in Python, R, JavaScript, Go, and Bash
- Implementing and deploying machine learning algorithms
- Setting up CI/CD pipelines with GitHub, Jenkins, and Netlify
- Analyzing data using R, Tidyverse, the PyData stack, and Spark