



Dr Alexander Engelhardt

Data Scientist

I am a data scientist specialized in Machine Learning. I combine a deep mathematical understanding of the relevant algorithms with the ability to communicate complex analyses clearly and practicably.

Version: August 6, 2019

Current Version available at http://alpha-epsilon.de/files/Profile_Engelhardt_en.pdf

Summary

| | |
|------------|--|
| Main Focus | Data Science Machine Learning, Deep Learning, tailor-made algorithms |
| | Programming R, Python, SQL, Linux shell |
| | Big Data Spark, Databricks, Amazon AWS, Microsoft Azure |
| | Optimizing Runtime optimization of programs and algorithms, automating complex workflows |
| | Communication Courses, user-friendly conveying of methods and results, technical and applied writing |
| Industries | Finance, energy trading, market research, bioinformatics, clinical research, start-ups, universities |
| Languages | fluent English and German |

Selected projects

My complete project history is enclosed below.

- 08/2018–heute **Full Stack Data Scientist, Energy Analytics, E.ON SE, Munich.**
– End-to-End Deployment of a prediction model for energy generation on Microsoft Azure
– Developed per-day predictive models for energy consumption of private households
– Developed models for anomaly detection in suspicious energy consumption
Tools used: Python, Spark, Azure Cloud, Docker, Databricks
- 12/2014–today **Lecturer, Munich R courses, department of statistics, LMU Munich, <http://www.muenchner-r-kurse.de/>.**
Delivered and developed the courses “Data Science theory”, “Practical data analysis with R”, and “Programming with R”
- 04/2018–07/2018 **Big Data Engineering, aifora GmbH, Düsseldorf.**
Developed an algorithm for pricing strategies in the retail textile industry. Implemented a data processing pipeline from raw data to an internal Hive storage, with Spark in Python and R.
Tools used: AWS, Spark, Hadoop, Python, Hive, Databricks

Konrad-Celtis-Str. 35, 2.OG – 81369 München

☎ +49 176 5690 6728 • ✉ engelhardt@alpha-epsilon.de • 🌐 www.alpha-epsilon.de

1/5

- 06/2016–07/2016 **Price Forward Curve for electricity prices**, *Bayerngas Energy GmbH, Munich*.
- 05/2017–03/2018 Three successive contracts:
- Created a program for automatic pricing of customer requests
 - Designed and programmed a real-time algorithm that generates price forward curves from market quotes.
 - Programmed a Shiny application for interactive tuning of parameters for a trading algorithm
 - Optimized the runtime of an R program for real-time pricing of trading data from 120 seconds per iteration to 15 seconds.
- Tools used: R, RStudio, Shiny, VBA
- 01/2017–06/2017 **Codevelopment of a product recommender**, *BASF SE, Ludwigshafen*.
- Implemented an algorithm to interpret interactions in xgboost models
 - Benchmarking against market basket analysis / association rules
- Tools used: R, xgboost
- 04/2015–06/2017 **Efficient parameter estimation in R**, *IBE, LMU Munich*.
- Designed and programmed a runtime-efficient EM algorithm to estimate risk parameters for cancer patients.
- Tools used: R, R with C++, parallel processing (BatchJobs package), cluster computing (Sun Grid Engine)

Community & Open Source

- mlr **Machine Learning in R**, [R package](#).
Contributor to the R package mlr. Various extensions and bugfixes as pull requests on GitHub.
- Deep Learning **Startup Name Generator**, [on GitHub](#).
Python package to generate name suggestions for companies and software.
- GenoGAM **A GAM based framework for analysis of ChIP-Seq data**, [Bioconductor package](#).
Co-development of an R package to compute parallelized statistical models on DNA.
- Blogging **Two Blogs on Statistics and Machine Learning**.
<http://www.crashkurs-statistik.de> - Statistics for non-statisticians (German)
<http://www.alpha-epsilon.de/blog> - All Things Data Science

Talks

- 10/2018 **Generate Company Names With Neural Networks.**, *Lightning Talk, PyConDE, 2018*.
A presentation of my Python package startup-name-generator
<https://www.youtube.com/watch?v=1w3Q3CE1dG0>

IT Skills

- Languages Python, R, SQL, Linux shell
- Machine Learning scikit-learn, Keras, mlr
- Big Data Spark, Databricks
- Cloud Computing Amazon Web Services (AWS), Microsoft Azure
- Misc Git, Docker, Emacs

Certifications / Awards

- 2018 Amazon Web Services (AWS) - Certified Solutions Architect - Associate
- 2018 Professional Scrum Master I, Scrum.org
- 2017 Second place, [IT-Freelancer of the year, 2017](#)
- 2017 Cloudera Certified Spark and Hadoop Developer
- 2013 Certificate of Proficiency in English, Grade A, University of Cambridge

Education

- 2013–2017 **PhD (Dr. rer. nat) Statistics**, *IBE, LMU Munich*, magna cum laude.
Dissertation: Efficient estimation algorithms for large and complex data sets
Available online at <https://edoc.ub.uni-muenchen.de/21020/>
- 2011–2013 **M.Sc. Statistics**, *Ludwig-Maximilian-University*, Munich.
- 2008–2011 **B.Sc. Statistics**, *Ludwig-Maximilian-University*, Munich.

Konrad-Celtis-Str. 35, 2.OG – 81369 München

☎ +49 176 5690 6728 • ✉ engelhardt@alpha-epsilon.de • 🌐 www.alpha-epsilon.de

2005–2008 **Vocational training as IT specialist – application development**, *Federal Office of Migration and Refugees*, Nuremberg.

Languages

German Native language
English Fluent (C2)
Spanish Basics (A2)

Online profiles

Website <http://www.alpha-epsilon.de>
GitHub <https://github.com/alexengelhardt>
StackOverflow <https://stackoverflow.com/users/477883/alexander-engelhardt>
XING https://www.xing.com/profile/Alexander_Engelhardt10
LinkedIn <https://www.linkedin.com/in/alexander-engelhardt-61b270a8>

August 6, 2019



Alexander Engelhardt

Enclosures: Complete project history; Publications

Complete project history

- 08/2018–heute **Full Stack Data Scientist, Energy Analytics, E.ON SE, Munich.**
– End-to-End Deployment of a prediction model for energy generation on Microsoft Azure
– Developed per-day predictive models for energy consumption of private households
– Developed models for anomaly detection in suspicious energy consumption
Tools used: Python, Spark, Azure Cloud, Docker, Databricks
- 04/2018–07/2018 **Big Data Engineering, aifora GmbH, Düsseldorf.**
Developed an algorithm for pricing strategies in the retail textile industry. Implemented a data processing pipeline from raw data to an internal Hive storage, with Spark in Python and R.
Tools used: AWS, Spark, Hadoop, Python, Hive, Databricks
- 04/2018–07/2018 **Predictive Analytics, Milon Care GmbH.**
Developed a model for the automatic computing of settings of exercise equipment based on body segment lengths
Tools used: Python, scikit-learn, Jupyter, Scrum/Agile
- 12/2014–today **Lecturer, Munich R courses, department of statistics, LMU Munich, <http://www.muenchner-r-kurse.de/>.**
Delivered and developed the courses “Data Science theory”, “Practical data analysis with R”, and “Programming with R”
- 01/2018–02/2018 **Search Engine Optimization, Artios.io, London.**
– Website audit with [sitespeed.io](https://www.sitespeed.io)
– Setup of an AWS instance for webcrawling with Sitebulb
– Co-development of an algorithm for automatic keyword generation for websites
Tools used: Amazon Web Services (EC2), R
- 06/2016–07/2016 **Price Forward Curve for electricity prices, Bayerngas Energy GmbH, Munich.**
05/2017–03/2018 Three successive contracts:
– Created a program for automatic pricing of customer requests
– Designed and programmed a real-time algorithm that generates price forward curves from market quotes.
– Programmed a Shiny application for interactive tuning of parameters for a trading algorithm
– Optimized the runtime of an R program for realtime-pricing of trading data from 120 seconds per iteration to 15 seconds.
Tools used: R, RStudio, Shiny, VBA
- 05/2017–10/2017 **Automatic error correction in time series, Trumpf AG, Ditzingen.**
Developed an algorithm that automatically finds and corrects wrong counter readings
Tools used: R
- 01/2017–06/2017 **Codevelopment of a product recommender, BASF SE, Ludwigshafen.**
– Implemented an algorithm to interpret interactions in xgboost models
– Benchmarking against market basket analysis / association rules
Tools used: R, xgboost
- 04/2015–06/2017 **Efficient parameter estimation in R, IBE, LMU Munich.**
Designed and programmed a runtime-efficient EM algorithm to estimate risk parameters for cancer patients.
Tools used: R, R with C++, parallel processing (BatchJobs package), cluster computing (Sun Grid Engine)
- 08/2015–01/2017 **Automating recurring analyses, Ipsos Loyalty GmbH.**
Developed scripts to automatically process and analyze quarterly recurring data.
Tools used: Python, SPSS (Macros)
- 10/2013–12/2015 **Parallelization of a statistical model, Gene center, LMU Munich.**
Parallelized a GAM (Generalized Additive Model) via a MapReduce approach in R.
Tools used: R, parallel processing (BatchJobs package)
- 11/2015–01/2016 **Directing a market research project, Sport- und Freizeitparadies GYM 80 GmbH.**
Developing the study design and subsequent statistical analysis and reporting.
Tools used: R
- 10/2015–12/2015 **Data analysis for a research project, University hospital, Tirol, Austria.**
Developed a statistical model for the prognosis of a biomarker based on various clinical parameters.
Tools used: Mixed linear model, R, dynamic reporting with RMarkdown

- 08/2015–09/2015 **Statistical analysis of an empirical study**, *Faculty of psychology and pedagogy, LMU Munich*.
Determining relevant factors for the success of early childhood education programs. In cooperation with the Kindernothilfe program.
Tools used: SPSS, mixed linear model
- 04/2013–09/2013 **Machine learning comparison study**, *IMBE, university of Erlangen*.
Planning and implementation of a comparison study of multiple machine learning algorithms to classify reflectance spectra.
Link to the publication: <http://dx.doi.org/doi:10.1186/1471-2288-14-91>
Tools used: R, cluster computing (Sun Grid Engine)
- 03/2012–05/2012 **Developing a prognostic model**, *Munich start-up*.
Forecasting user numbers of a smartphone application to win investors by combining demographic data and statistical growth models.
Tools used: R

Employment

- 03/2013–09/2013 **Research Assistant**, *Statistical consulting unit (StaBLab)*, LMU Munich.
Statistical consulting of students and externals
- 05/2011–04/2013 **Working student**, *STAT-UP Statistical Consulting & Services*, Munich.
Programming statistical solutions in R and SPSS
Relevant projects:
 - Developed R packages, scripts and documentation for food safety for the Federal Institute of Risk Assessment (BfR)
 - Developed a database (MySQL) in a project to compute growth and inactivation parameters of microbiological models

Publications

In magazines

- Schnelleinstieg in Data Science und Machine Learning**. Appeared in *IT Freelancer Magazin*, 09/2017. Available at <http://www.it-freelancer-magazin.de/index.php/2017/09/06/schnelleinstieg-in-data-science-und-machine-learning/>
- Eine Einführung ins Machine Learning**. Appeared in *VisualStudio1.de*, Issue 04/2015. Available at <http://www.alpha-epsilon.de/files/vs1-MachineLearning.pdf>
- Im Sog der Daten – Big Data Analytics mit Revolution R** Appeared in *VisualStudio1.de*, Issue 03/2015. Available at <http://www.alpha-epsilon.de/files/vs1-BigData.pdf>

Scientific papers

- Engelhardt *et al.*: **Efficient Maximum Likelihood Estimation for Pedigree Data with the Sum-Product Algorithm**. *Human Heredity*, 2017
<https://doi.org/10.1159/000475465>
- Stricker, Engelhardt, *et al.*: **GenoGAM: Genome-Wide Generalized Additive Models for ChIP-Seq Analysis**. *Bioinformatics*, 2017
<https://doi.org/10.1093/bioinformatics/btx150>
- Engelhardt *et al.*: **Constructing an ROC Curve to Assess a Treatment-Predictive Continuous Biomarker**. *Studies in health technology and informatics*, 2016
<http://dx.doi.org/doi:10.3233/978-1-61499-678-1-745>
- Engelhardt *et al.*: **Comparing classification methods for diffuse reflectance spectra to improve tissue specific laser surgery**. *BMC Medical Research Methodology*, 2014
<http://dx.doi.org/doi:10.1186/1471-2288-14-91>