



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: September 3, 2023

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 Python, R, 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.

- 11/2022–03/2023 **Data Scientist, Shape Risk Management, E.ON SE, Essen.**
– Development of browser-based interactive risk reports for shape and performance of a energy PFC
Tools used: Python, streamlit, pandas, Azure DevOps
- 05/2020–10/2022 **Data Engineer, Energy Trading, EnBW AG, Karlsruhe.**
– Engineering of a data processing pipeline that computes Profit and Loss for energy contracts
– Migration of an automatic data processing system from Palantir Foundry to Amazon Web Services
– Creation of analyses and reports for users in the trading team
Tools used: Palantir Foundry, AWS (Lambda, Glue, S3, SageMaker), Python, pySpark, SQL, Azure DevOps
- 08/2018–03/2020 **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 a Python package for predictive models for energy consumption of private households
– Developed models for anomaly detection of suspicious energy consumption
Tools used: Python, xgboost, Spark, SQL, Azure Cloud, Docker, Databricks
- 12/2014–today **Lecturer, Essential Data Science GmbH, Munich, <https://www.essentialds.de/>.**
Delivered and developed the courses “Data Science theory”, “Practical data analysis with R”, and “Programming with R”

Sylvensteinstraße 2 – 81369 München

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

1/6

- 04/2018–07/2018 **Big Data Engineering**, *aifora GmbH, Düsseldorf*.
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, SQL, Hive, Databricks
- 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

- 10/2019 **Interpretable Machine Learning**, *Presentation*, PyConDE Berlin, 2019.
How to make black box models explainable
<https://www.youtube.com/watch?v=sAqSGY-HkVY>
- 10/2019 **Adversarial Machine Learning**, *Lightning Talk*, PyConDE Berlin, 2019.
How to attack a machine learning model
https://www.youtube.com/watch?v=PZWqFa_8AEU
- 10/2018 **Generate Company Names With Neural Networks**, *Lightning Talk*, PyConDE Karlsruhe, 2018.
A presentation of my Python package startup-name-generator
<https://www.youtube.com/watch?v=1w3Q3CEldG0>

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

Publications

Books

Crashkurs Statistik. 2020. Available at <https://www.amazon.de/dp/B086G17BDM>

In magazines

Sylvensteinstraße 2 – 81369 München

Künstliche Intelligenz interpretierbar machen. Appeared in *Informatik Aktuell*, 11/2019. Available at <https://www.informatik-aktuell.de/betrieb/kuenstliche-intelligenz/kuenstliche-intelligenz-interpretierbar-machen.html>

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>

Certifications / Awards

- 2023 [Microsoft Certified: Azure Developer Associate](#)
- 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.
- 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

Sylvensteinstraße 2 – 81369 München

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

References

Dr. Alexander Engelhardt, within our team, contributed at improving the accuracy of short-term renewable generation forecasts. He was responsible for the design and implementation of cloud-based end-to-end data processing and machine learning pipelines, exploiting actual asset production information as well as data from external providers and weather services. He worked in a highly proactive manner and completely identified with his tasks and our company at all times. He always displayed an impressive level of dedication and motivation.

We would like to thank him for his consistently very good performance and wish him all the best and much success in his professional and personal future.

– Dr. Giorgio Cortiana, Head of Advanced Analytics Energy Intelligence at E.ON, April 2020

Alexander war bei uns als R-Entwickler tätig und hat ein Programm zur automatischen Bepreisung von Strom- und Gasprodukten weiterentwickelt und bezüglich der Laufzeit optimiert.

Wir haben Alexander als kreativen und lösungsorientierten Kopf kennengelernt, der auch in hektischen Phasen fokussiert bleibt und das Ziel nicht aus den Augen verliert. Er hat sein Wissen stets bereitwillig an mich und meine Mitarbeiter weitergegeben, sowohl während dem Tagesgeschäft als auch in Form einer inhouse-Schulung. Mit dem Ergebnis seiner Arbeit sind wir sehr zufrieden.

– Martin Baier, Abteilungsleiter Pricing, Bayerngas Energy GmbH, im März 2018

Herr Engelhardt hat uns während einer Hochphase im Projekt mit großem zeitlichen Druck sehr unterstützt. Beeindruckend war vor allem, wie schnell und tief er sich in die komplexen Inhalte und Zusammenhänge eingearbeitet hat. Mit seinem statistischen Wissen und Ideen zur Effizienzsteigerung war er für unser Team sehr wertvoll. Da war es aus unserer Sicht nur logisch, dass wir im Anschluss an die Projektarbeit uns von Herrn Engelhardt nur zu gerne im Rahmen einer Schulung sein umfangreiches Wissen vermitteln lassen wollten. Auch hier haben uns seine Unterstützung und Ratschläge sehr viel weiter gebracht. Wir danken Herrn Engelhardt für die gemeinsame Zeit und kommen bei Bedarf gerne wieder auf ihn zu.

– Martina Lorenz, Senior Research Executive, Ipsos Loyalty GmbH, im September 2016

Zuverlässig, kompetent und extrem hilfreich war Herr Engelhardt bei der statistischen Auswertung einer sehr komplex angelegten international-vergleichenden Studie zur frühkindlichen Bildung. Ohne seine Expertise hätten wir den ohnehin sehr ambitionierten Zeitplan nicht einhalten können. Beeindruckt hat mich als Forschungsleiter wie rasch und tief Herr Engelhardt den sozialwissenschaftlichen Forschungsgegenstand hat durchdringen können, um die für uns relevanten Daten zu rechnen, sie als Ergebnisse darstellen und datenbasiert interpretieren zu können.

– Prof. Dr. Reinhard Markowitz, LMU München, im März 2016

September 3, 2023



Alexander Engelhardt

Enclosures: Complete project history; Publications

Complete project history

- 11/2022–03/2023 **Data Scientist, Shape Risk Management, E.ON SE, Essen.**
– Development of browser-based interactive risk reports for shape and performance of a energy PFC
Tools used: Python, streamlit, pandas, Azure DevOps
- 05/2020–10/2022 **Data Engineer, Energy Trading, EnBW AG, Karlsruhe.**
– Engineering of a data processing pipeline that computes Profit and Loss for energy contracts
– Migration of an automatic data processing system from Palantir Foundry to Amazon Web Services
– Creation of analyses and reports for users in the trading team
Tools used: Palantir Foundry, AWS (Lambda, Glue, S3, SageMaker), Python, pySpark, SQL, Azure DevOps
- 08/2018–03/2020 **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 a Python package for predictive models for energy consumption of private households
– Developed models for anomaly detection of suspicious energy consumption
Tools used: Python, xgboost, Spark, SQL, Azure Cloud, Docker, Databricks
- 04/2018–07/2018 **Big Data Engineering, aifora GmbH, Düsseldorf.**
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, SQL, 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, Essential Data Science GmbH, Munich, <https://www.essentialds.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 Kindernotheilfe 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