



Dr. Alexander Engelhardt

Data Scientist

Ich bin Data Scientist im Bereich Machine Learning, und verbinde ein tiefes mathematisches Verständnis der relevanten Algorithmen mit der Fähigkeit, komplexe Analysen in klarer und umsetzbarer Form zu kommunizieren.

Stand: 6. August 2019

Aktuelle Version verfügbar unter <http://alpha-epsilon.de/profil/>

Zusammenfassung

Schwerpunkte **Data Science**

Machine Learning, Deep Learning, maßgeschneiderte Algorithmen

Programmieren

R, Python, SQL, Linux shell

Big Data

Spark, Databricks, Amazon AWS, Microsoft Azure

Optimierung

Laufzeitoptimierung von Programmen und Algorithmen, Automatisierung von komplexen Arbeitsabläufen

Kommunikation

Schulungen, anwenderfreundliches Vermitteln von Methoden und Ergebnissen, technisches sowie angewandtes Schreiben

Branchenerfahrung Finanzen, Energie, Marktforschung, Bioinformatik, klinische Forschung, Start-Ups, Universitäten

Sprachen fließend Deutsch und Englisch

Ausgewählte Projekte

Meine vollständige Projekthistorie finden Sie im Anhang.

08/2018–heute **Full Stack Data Scientist, Energy Analytics, E.ON SE, München.**

– End-to-End Deployment eines Prognosemodells für Stromgeneration auf Microsoft Azure

– Entwickelte tagesgenaue Prognosemodelle für Energieverbrauch von Privathaushalten

– Entwickelte Modelle zur Anomalieerkennung von ungewöhnlichem Stromverbrauch

Verwendete Tools: Python, Spark, Azure Cloud, Docker, Databricks

12/2014–heute **Dozent, Münchner R-Kurse, Institut für Statistik der LMU, München, <http://www.muenchner-r-kurse.de/>.**

Halten und Weiterentwickeln der Kurse "Data Science Methodenkurs", "Praktische Datenanalyse mit R", und "Programmieren mit R"

04/2018–07/2018 **Big Data Engineering, aifora GmbH, Düsseldorf.**

Entwickelte einen Algorithmus für Pricing-Strategien im Textileinzelhandel.

Implementierte eine Datenverarbeitungs-Pipeline, von gelieferten Daten in einen internen Hive-Speicher, mit Spark in Python und R.

Verwendete Tools: AWS, Spark, Hadoop, Python, Hive, Databricks

Konrad-Celtis-Str. 35 – 81369 München

☎ +49 (0)176 5690 6728 • ✉ engelhardt@alpha-epsilon.de

🌐 www.alpha-epsilon.de

- 06/2016–07/2016 **Price Forward Curve für Strompreise**, *Bayerngas Energy GmbH, München*.
 05/2017–03/2018 Drei Folgeaufträge:
 – Erstellte ein Programm zum automatischen Pricing von Kundenanfragen
 – Entwarf und programmierte einen Algorithmus zur Generierung von Price Forward Curves für Strompreise, von Stunden- bis Monatsgranularität
 – Programmierte eine Shiny-Applikation zur explorativen Einstellung von Parametern für einen Handelsalgorithmus
 – Optimierte die Laufzeit eines R-Programms zum realtime-Pricing von Handelsdaten von 120 Sekunden pro Iteration auf 15 Sekunden
 Verwendete Tools: R, RStudio, Shiny, VBA
- 01/2017–06/2017 **Mitentwicklung an einem Produktempfehlungssystem**, *BASF, Ludwigshafen*.
 – Implementierte einen Algorithmus zur Interpretation von Interaktionen in xgboost-Modellen
 – Benchmarking gegen Warenkorbanalyse / Assoziationsregeln
 Verwendete Tools: R, xgboost
- 04/2015–06/2017 **Effiziente Parameterschätzung in R**, *IBE, LMU München*.
 Entwarf und programmierte einen lauffzeiteffizienten EM-Algorithmus zur Schätzung von Risikoparametern bei Krebspatienten.
 Verwendete Tools: R, R mit C++, Parallel Processing (BatchJobs-Paket), Cluster computing (Sun Grid Engine)

Community & Open-Source

- mlr **Machine Learning in R**, *R-Paket*.
 Contributor des R-Pakets mlr. Diverse Erweiterungen und Bugfixes als Pull-Requests auf GitHub.
- Deep Learning **Startup Name Generator**, *auf GitHub*.
 Python-Paket zum Generieren von Namensvorschlägen für Unternehmen und Software.
- GenoGAM **A GAM based framework for analysis of ChIP-Seq data**, *Bioconductor-Paket*.
 Miterstellung eines R-Pakets zum Berechnen parallelisierter statistischer Modelle auf DNA.
- Blogging **Zwei Blogs über Statistik und Machine Learning**.
<http://www.crashkurs-statistik.de> - Statistik für Nicht-Statistiker
<http://www.alpha-epsilon.de/blog> - All Things Data Science

Vorträge

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

IT-Kenntnisse

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

Zertifikate / Auszeichnungen

- 2018 Amazon Web Services (AWS) - Certified Solutions Architect - Associate
 2018 Professional Scrum Master I, Scrum.org
 2017 2. Platz, *IT-Freelancer des Jahres 2017*
 2017 Cloudera Certified Spark and Hadoop Developer
 2013 Certificate of Proficiency in English, Grade A, University of Cambridge

Ausbildung

- 2013–2017 **Promotion (Dr. rer. nat)**, *IBE, Ludwig-Maximilians-Universität München*, magna cum laude.
 Dissertation: Efficient estimation algorithms for large and complex data sets
 Online verfügbar unter <https://edoc.ub.uni-muenchen.de/21020/>
- 2011–2013 **Master-Studium der Statistik**, *Ludwig-Maximilians-Universität, München*.

Konrad-Celtis-Str. 35 – 81369 München

☎ +49 (0)176 5690 6728 • ✉ engelhardt@alpha-epsilon.de
 🌐 www.alpha-epsilon.de

2008–2011 **Bachelor-Studium der Statistik**, *Ludwig-Maximilians-Universität*, München.

2005–2008 **Ausbildung zum Fachinformatiker – Anwendungsentwicklung**, *Bundesamt für Migration und Flüchtlinge*, Nürnberg.

Sprachen

Deutsch Muttersprache
Englisch Fließend (C2)
Spanisch Grundkenntnisse (A2)

Onlineprofile

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>

Kundenstimmen

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 Markowetz, LMU München, im März 2016

6. August 2019



Alexander Engelhardt

Anlagen: Vollständige Projekthistorie; Veröffentlichungen

Konrad-Celtis-Str. 35 – 81369 München

☎ +49 (0)176 5690 6728 • ✉ engelhardt@alpha-epsilon.de

🌐 www.alpha-epsilon.de

Vollständige Projekthistorie

- 08/2018–heute **Full Stack Data Scientist, Energy Analytics, E.ON SE, München.**
– End-to-End Deployment eines Prognosemodells für Stromgeneration auf Microsoft Azure
– Entwickelte tagesgenaue Prognosemodelle für Energieverbrauch von Privathaushalten
– Entwickelte Modelle zur Anomalieerkennung von ungewöhnlichem Stromverbrauch
Verwendete Tools: Python, Spark, Azure Cloud, Docker, Databricks
- 04/2018–07/2018 **Big Data Engineering, aifora GmbH, Düsseldorf.**
Entwickelte einen Algorithmus für Pricing-Strategien im Textileinzelhandel.
Implementierte eine Datenverarbeitungs-Pipeline, von gelieferten Daten in einen internen Hive-Speicher, mit Spark in Python und R.
Verwendete Tools: AWS, Spark, Hadoop, Python, Hive, Databricks
- 04/2018–07/2018 **Predictive Analytics, Milon Care GmbH.**
Programmierte ein Modell zur automatischen Berechnung von Einstellungen von Fitnessgeräten aufgrund von Körpersegmentlängen
Verwendete Tools: Python, scikit-learn, Jupyter, Scrum/Agile
- 12/2014–heute **Dozent, Münchner R-Kurse, Institut für Statistik der LMU, München, <http://www.muenchner-r-kurse.de/>.**
Halten und Weiterentwickeln der Kurse "Data Science Methodenkurs", "Praktische Datenanalyse mit R", und "Programmieren mit R"
- 01/2018–02/2018 **Search Engine Optimization, Artios.io, London.**
– Webseiten-Audit mit sitespeed.io
– Setup einer AWS-Instanz zum Web-crawling mit Sitebulb
– Mitarbeit an einem Algorithmus zum automatischen Vorschlagen von Keywords für Webseiten
Verwendete Tools: Amazon Web Services (EC2), R
- 06/2016–07/2016 **Price Forward Curve für Strompreise, Bayerngas Energy GmbH, München.**
05/2017–03/2018 Drei Folgeaufträge:
– Erstellte ein Programm zum automatischen Pricing von Kundenanfragen
– Entwarf und programmierte einen Algorithmus zur Generierung von Price Forward Curves für Strompreise, von Stunden- bis Monatsgranularität
– Programmierte eine Shiny-Applikation zur explorativen Einstellung von Parametern für einen Handelsalgorithmus
– Optimierte die Laufzeit eines R-Programms zum realtime-Pricing von Handelsdaten von 120 Sekunden pro Iteration auf 15 Sekunden
Verwendete Tools: R, RStudio, Shiny, VBA
- 05/2017–10/2017 **Automatische Fehlerkorrektur in Zeitreihen, Trumpf AG, Ditzingen.**
Entwickelte einen Algorithmus, der automatisch falsche Werte in Zählerständen erkennt und korrigiert
Verwendete Tools: R
- 01/2017–06/2017 **Mitentwicklung an einem Produktempfehlungssystem, BASF, Ludwigshafen.**
– Implementierte einen Algorithmus zur Interpretation von Interaktionen in xgboost-Modellen
– Benchmarking gegen Warenkorbanalyse / Assoziationsregeln
Verwendete Tools: R, xgboost
- 04/2015–06/2017 **Effiziente Parameterschätzung in R, IBE, LMU München.**
Entwarf und programmierte einen laufzeiteffizienten EM-Algorithmus zur Schätzung von Risikoparametern bei Krebspatienten.
Verwendete Tools: R, R mit C++, Parallel Processing (BatchJobs-Paket), Cluster computing (Sun Grid Engine)
- 08/2015–01/2017 **Automatisierung wiederkehrender Analysen, Ipsos Loyalty GmbH.**
Entwickelte Skripte zur automatisierten Auswertung quartalsweise wiederkehrender Datenauswertungen.
Verwendete Tools: Python, SPSS (Makros)
- 10/2013–12/2015 **Parallelisierung eines statistischen Modells, Genzentrum, LMU München.**
Parallelisierte via MapReduce-Ansatz ein GAMs (Generalisiertes Additives Modell) in R.
Verwendete Tools: R, Parallel Processing (BatchJobs-Paket)
- 11/2015–01/2016 **Leitung eines Marktforschungsprojekts, Sport- und Freizeitparadies GYM 80 GmbH.**
Leitung eines Marktforschungsprojekts. Entwurf des Untersuchungsdesigns und anschließende statistische Auswertung und Berichterstellung.
Verwendete Tools: R

- 10/2015–12/2015 **Datenanalyse für eine Forschungsarbeit**, *Universitätsklinikum, Tirol*.
Entwickelte ein statistisches Prognosemodell für einen Biomarker anhand verschiedener klinischer Parameter.
Verwendete Tools: Gemischtes lineares Modell mit R, dynamische Berichterstellung mit RMarkdown
- 08/2015–09/2015 **Statistische Auswertung einer empirischen Studie**, *Fakultät für Psychologie und Pädagogik, LMU München*.
Bestimmung relevanter Faktoren auf den Erfolg von frühkindlichen Bildungsprogrammen. In Zusammenarbeit mit der Kindernothilfe.
Verwendete Tools: SPSS, Gemischtes Lineares Modell
- 04/2013–09/2013 **Machine Learning-Vergleichsstudie**, *IMBE, FAU Erlangen*.
Planung und Durchführung einer Vergleichsstudie von Machine Learning-Algorithmen zur Klassifikation von Reflektanzspektren.
Link zur Veröffentlichung: <http://dx.doi.org/doi:10.1186/1471-2288-14-91>
Verwendete Tools: R, Cluster Computing (Sun Grid Engine)
- 03/2012–05/2012 **Entwicklung eines Prognosemodells**, *Münchner Start-Up*.
Prognose der Benutzerzahlen einer Smartphone-App zur Investorengewinnung mit Hilfe der Verbindung von demographischen Daten und statistischen Wachstumsmodellen.
Verwendete Tools: R

Angestellt

- 03/2013–09/2013 **Wissenschaftliche Hilfskraft**, *Statistisches Beratungslabor (StaBLab) der LMU, München*.
Statistische Beratung von Studierenden und Externen
- 05/2011–04/2013 **Werkstudent**, *STAT-UP Statistical Consulting & Services, München*.
Programmierung statistischer Lösungen in R und SPSS
Relevante Projekte:
 - Entwickelte R-Pakete, -Skripte und Dokumentationen im Bereich Lebensmittelsicherheit für das Bundesinstitut für Risikobewertung (BfR)
 - Entwarf eine Datenbank (MySQL) in einem Projekt zur Berechnung von Wachstums- und Inaktivierungsparametern von mikrobiologischen Modellen

Veröffentlichungen

In Magazinen

Schnelleinstieg in Data Science und Machine Learning. Erschienen im *IT Freelancer Magazin*, 09/2017. Verfügbar unter <http://www.it-freelancer-magazin.de/index.php/2017/09/06/schnelleinstieg-in-data-science-und-machine-learning/>

Eine Einführung ins Machine Learning. Erschienen in *VisualStudio1.de*, Ausgabe 04/2015. Verfügbar unter <http://www.alpha-epsilon.de/files/vs1-MachineLearning.pdf>

Im Sog der Daten – Big Data Analytics mit Revolution R Erschienen in *VisualStudio1.de*, Ausgabe 03/2015. Verfügbar unter <http://www.alpha-epsilon.de/files/vs1-BigData.pdf>

Wissenschaftliche Artikel

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>