

Sebastian Jäger

████████████████████ Berlin – Germany
☎ (+49) ██████████ • ✉ mail@sebastian-jaeger.me
🌐 sebastian-jaeger.me • 🌐 GitHub • in LinkedIn
🔍 Google Scholar



PhD student interested in data integration and quality with several years of experience of working in and leading research and customer projects. Experienced with common languages, libraries, tools, cluster, and cloud infrastructures. Always eager to gain new and share knowledge, apply cutting edge technologies, and give talks.

Education/Courses

PhD Student **Berlin University of Applied Sciences and Technology (BHT), Germany**
Machine Learning *since Nov 2020*

Research Project: Green Consumption Assistant ([link](#))

Research Interests: Data Quality, Data Imputation, Data Cleaning

Data Science **Berlin University of Applied Sciences and Technology (BHT), Germany**
Master of Science *Oct 2018 - Oct 2020*
Final grade: 1.0

Thesis Title: Compressing BERT – An Evaluation and Combination of Methods

We analyzed several compressing approaches for BERT and applied them on models for the medical domain.

Supervisors: Prof. Dr.-Ing. habil. Alexander Löser & Prof. Dr. rer. nat. Felix Bießmann

Deep Learning Specialization **Massive Open Online Course (MOOC), Online**
Coursera Course *Apr 2018*

Covers the foundations of Deep Learning, the structure of machine learning projects, CNNs, RNNs, as well as optimization, initialization, and regularization techniques.

Computer Science **Karlsruhe University of Applied Sciences, Germany**
Bachelor of Science *Oct 2014 – Feb 2018*
Final grade: 1.2

Thesis Title: Horizontales Skalieren von Deep Learning Frameworks

We evaluated the horizontal scaling properties of two common Deep Learning Frameworks experimentally, namely TensorFlow and MXNet.

Supervisors: Prof. Dr. rer. nat. Christian Zirpins & Hans-Peter Zorn (inovex GmbH)

Work Experience

Berlin University of Applied Sciences and Technology (BHT) **Berlin, Germany**
PhD Student *since Nov 2020*

Data Science and Text-based Information Systems (DATEXIS) **Berlin, Germany**
Thesis Student *Mar 2020 - Oct 2020*

inovex GmbH
Working Student

Karlsruhe, Germany
Aug 2017 – Feb 2020

Kubernetes-based data science platform: Internal used platform to speedup and scale data science projects.

Responsibilities:

- Implementation and security of the Kubernetes cluster environment
- Selection and combination of components and tools
- Implementation of a Go-based CLI tool to easily interact with the platform

Metadata management system for 3D mass spectroscopy data: Research project for analysis of 3D mass spectroscopy data. In collaboration with the Mannheim University of Applied Sciences.

Responsibilities:

- Leading and coordinating a team of five students
- Communicating and presenting the project process internally and to public

exensio GmbH
Internship & Working Student

Karlsruhe, Germany
Mar 2016 – May 2017

- Java and Groovy based software development
- Move internal infrastructure to a Docker-based environment

Publications

Conference Paper
AISTATS

Valencia, Spain
May 2024

Title: From Data Imputation to Data Cleaning – Automated Cleaning of Tabular Data Improves Downstream Predictive Performance

Workshop Paper
DataPerf

Baltimore, USA
Jul 2022

Title: GreenDB - A Dataset and Benchmark for Extraction of Sustainability Information of Consumer Goods ([DOI](#))

Preprint
arXiv

Online
May 2022

Title: GreenDB: Toward a Product-by-Product Sustainability Database ([DOI](#))

Journal Article
Frontiers in Sustainability

Online
Jul 2022

Title: Nudging Sustainable Consumption: A Large-Scale Data Analysis of Sustainability Labels for Fashion in German Online Retail ([DOI](#))

Journal Article
Frontiers in Big Data

Online
Jul 2021

Title: A Benchmark for Data Imputation Methods ([DOI](#))

Conference Talk
data2day

Ludwigshafen, Germany
Oct 2019

Title: Mit Metadatenmanagement hinzu reproduzierbaren und flexiblen Data-Science-Workflows auf Kubernetes ([link](#))

Workshop Paper*Distributed Infrastructures for Deep Learning***Rennes, France***Dec 2018***Title:** Parallelized Training of Deep NN – Comparison of Current Concepts and Frameworks ([DOI](#))**Article***heise Developer***Online***Nov 2018***Title:** Machine Learning im Kubernetes Cluster ([link](#))**Conference Talk***Minds Mastering Machines***Köln, Germany***Apr 2018***Title:** Skalieren Von Deep Learning Frameworks mit Hilfe von Cloudinfrastrukturen und Kubernetes ([link](#))