

# David Schulte

## EDUCATION

---

|                                                                       |                                 |
|-----------------------------------------------------------------------|---------------------------------|
| <b>Machine Learning (PhD)</b><br><i>Humboldt University of Berlin</i> | Berlin, DE<br>8/2026 – present  |
| <b>Statistics (M.Sc)</b><br><i>Humboldt University of Berlin</i>      | Berlin, DE<br>10/2020 – 3/2024  |
| <b>Computer Science (Semester abroad)</b><br><i>Waseda University</i> | Tokyo, JP<br>4/2018 – 8/2018    |
| <b>Mathematics (B.Sc)</b><br><i>University of Cologne</i>             | Cologne, DE<br>10/2014 – 9/2019 |

## WORK EXPERIENCE

---

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |                               |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------|
| <b>Machine Learning Engineer</b><br><i>Meltwater</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | Berlin, DE<br>9/2025 – 7/2026 |
| <ul style="list-style-type: none"><li>Built machine learning systems for Natural Language Processing<ul style="list-style-type: none"><li>Conceptualized efficient solutions for high-volume text classification</li><li>Curated and augmenting training data</li><li>Deployed models in production and monitored them</li></ul></li><li>Maintained services for generative language models</li></ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |                               |
| <b>Data Scientist</b><br><i>Vaayu</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | Berlin, DE<br>9/2022 – 8/2025 |
| <ul style="list-style-type: none"><li>Owned an NLP model that predicts ~1M product taxonomies per day<ul style="list-style-type: none"><li>Improved the architecture of the model for efficient hierarchical classification</li><li>Formulated a loss function to align model behavior and measure performance</li><li>Devised a routine for data collection to speed up training iterations</li><li>Maintained the production service to guarantee efficiency and accuracy</li></ul></li><li>Conducted large-scale analysis of retail shipping behavior and trends<ul style="list-style-type: none"><li>Built data pipelines for data aggregation and analysis</li><li>Conducted analysis of shipping behavior and communicated it to the CTO and clients</li><li>Built dashboards for interactive data exploration</li><li>Simulated the impact of reduction measures on logistics carbon emissions using a self-devised sampling strategy</li></ul></li><li>Built a multi-agent system to analyze semi-structured customer data</li><li>Consulted the CTO, the Head of Data, and clients about statistics-/ML-related topics</li></ul> |                               |
| <b>Data Scientist (Working Student)</b><br><i>Jobufo</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | Berlin, DE<br>3/2022 – 8/2022 |
| <ul style="list-style-type: none"><li>Developed a sequence labeling model for segmenting job advertisements</li><li>Used unsupervised clustering of webpages to facilitate market research</li></ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |                               |

## PUBLICATIONS

---

Less is More: Parameter-Efficient Selection of Intermediate Tasks for Transfer Learning      EMNLP 2024  
(Schulte et al., EMNLP 2024)

## PROJECTS

---

### hf-dataset-selector

A Python package for efficient selection of datasets for intermediate task transfer learning

### word-guesser

Solving a word guessing game by navigating through embeddings in a vector database

### Forecasting of German Solar Power Generetaion

Using a time series model to analyze and forecast solar power generation in Germany

### Reinforcement Learning for 3-Player Chinese Checkers

Designed and trained an agent to play a multiplayer board game that surpassed human skill-level

## SKILLS

---

**Concepts:**            Machine Learning, NLP, Statistics, Model Deployment, Data Analysis  
**Technologies:**    Python, PyTorch, SQL, Git, Docker, Kubernetes, Terraform, GCP, AWS  
**Languages:**        English (Fluent), German (Native), Japanese (Intermediate / JLPT N2)