

David Schulte

WORK EXPERIENCE

Data Scientist

Berlin, DE

Vaayu

9/2022 – present

- Owning an NLP model that predicts ~1M product taxonomies per day
- Conducting large-scale analysis of retail shipping behavior and trends
- Simulating the impact of reduction measures on logistics carbon emissions

Data Scientist (Working Student)

Berlin, DE

Jobufo

3/2022 – 8/2022

- Developed a sequence labeling model for segmenting job advertisements
- Used unsupervised clustering of webpages to facilitate market research

EDUCATION

Statistics (M.Sc)

Berlin, DE

Humboldt University of Berlin

10/2020 – 3/2024

Specialization in Data Science and Machine Learning

Mathematics (B.Sc)

Cologne, DE

University of Cologne

10/2014 – 9/2019

Courses in Probability Theory, Numerical Simulation and Computer Science

Computer Science (Semester abroad)

Tokyo, JP

Waseda University

4/2018 – 8/2018

Courses in Machine Learning, Data Mining and Information Security

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

Prediction of Customer Returns

Created and evaluated a model that predicts returns to an online retailer to minimize return costs

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, Docker, Kubernetes, Git, GCP

Languages: English (Fluent), German (Native), Japanese (Intermediate / JLPT N2)