Dr. Franziska Horn

hey@franziskahorn.de | linkedIn/franziska-horn | github/cod3licious | franziskahorn.de

A technical leader with a strong product mindset and a background in data science and software engineering with **11+ years of experience building data products** in both research and application contexts.

WORK EXPERIENCE

Freelance Data Science Solutions Architect | Leipzig (Germany) / remote 10/2018 - present
- strategy & ideation workshops with department heads and product managers to identify potential AI use cases
- design, implementation, and evaluation of data science solutions tailored to my client's needs (using Python)
- 1:1 coaching sessions and multi-day trainings on how (and when) to use machine learning techniques in practice
- clients include BASF and TRUMPF

Data & Solutions Engineering Team Lead | alcemy, Berlin (remote) 10/2022 - 11/2023
- hired and led a team of three data scientists and engineers, responsible for customer integration and data-related customer requests
- created team vision & strategy and prioritized projects & objectives together with the customer success, product, and sales teams
- designed & built web apps empowering the customer success team to conduct recurring customer data analyses independently
- analyzed, documented, and optimized internal processes, e.g., reduced onboarding time for new customers from 50 to 25 days

Senior Customer-Facing Data Scientist | alcemy, Berlin (remote) 02/2022 - 10/2022
- analyzed laboratory data from our customers (cement plants), e.g., to identify irregularities in their production processes
- reduced time spent on recurring analyses by implementing configurable, reusable report templates
- simplified data integration process for new customers by planning and implementing refactorings of core product components, minimized required code files per customer to 6 from 17

Postdoc Visiting Scientist | ML Group, TU (Technical University) Berlin (remote) 05/2020 - 11/2020
- developed continuously evolving word embeddings that account for meaning changes over time (published at ACL 2021)
- collaboration with Prof. Dr. Alan Akbik from the Humboldt University of Berlin

Data Science Consultant | BASF, Ludwigshafen 03/2018 - 10/2018
- most experienced Python developer and data scientist in the team, responsible for mentoring and code reviews
- implemented machine learning algorithms, e.g., the autofeat library for automated feature engineering and selection, and analyzed complex datasets to optimize processes in chemical plants

Research Assistant | ML Group, TU Berlin 09/2017 - 02/2018
- predictive maintenance / time series analysis project in collaboration with BASF (worked on-site in Ludwigshafen)
- designed, implemented, and evaluated linear and non-linear regression models in Python to predict the degradation of catalysts in chemical plants (patented and published in Computers and Chemical Engineering)

Machine Learning Team Lead | Spectrm, Berlin 07/2016 - 06/2017
- established the machine learning team and hired two ML engineers
- implemented a content recommendation API for newspaper articles to promote our clients' content (Python Flask App)
- developed a chatbot "AI" to respond to user messages automatically (using RiveScript)

Data Scientist | idalab, Berlin 02/2014 - 06/2016
- advanced analytics consulting projects, ML algorithm development in Python, presentation of results, and project management
- clients included razorfish (NLP backend for automatic content classification) and outfittery (style prediction for curated shopping)

Student Research Assistant | ML Group, TU Berlin 09/2013 - 09/2014
- focus on text classification, unsupervised learning (word2vec embeddings, dimensionality reduction), and information extraction
- short-term research stay at UCLA; collaboration with Prof. Dr. Alcino Silva

Student Research Assistant | BBCI Lab, TU Berlin 08/2012 - 08/2013
- EEG data analysis at the Berlin Brain-Computer Interface Lab; developed and efficiently implemented new algorithms in MatLab
- three peer-reviewed publications (journal and conferences)

Research Intern | MIT (Massachusetts Institute of Technology), Cambridge, MA 07/2011 - 10/2011
- at the McGovern Institute for Brain Research / Gabrieli Lab; analyzed fMRI data using Python (published in JAMA Psychiatry)
- sponsored by a DAAD RISE scholarship

Student Research Assistant | Fraunhofer Institute for Chemical Technology, Pfinztal 07/2007 - 12/2009
- worked independently, responsible for collection of infrared spectroscopy data

SKILLS

Leadership

I'm a "get sh*t done (well)" person, who motivates empowered teams through a strong vision and clear priorities, while striving for operational excellence in an agile environment.

- mentored junior team members and students and conducted individual data science coaching sessions as a freelancer
- gathered requirements from external customers and aligned team objectives with customer success, sales, and product teams
- managed the product backlog and held sprint planning meetings, while ensuring the tasks present growth opportunities for individuals
- facilitated workshops and architectural decision making processes across multiple teams

Certificates: Professional Scrum Master 1 & Facilitation Skills (Scrum.org)

Software Engineering

Crafting decoupled software architectures aligned with business domains and value streams, while enjoying the occasional refactoring.

- designed and implemented various web apps (e.g., data tools at alcemy, prototypes for freelance clients, and personal projects)
- authored open source Python libraries, e.g., autofeat for automated feature engineering with 450+ stars on GitHub
- worked on a shared codebase, incl. contributing to production code and tests, and reviewing merge requests

Tools: Python (incl. poetry, ruff, mypy, FastAPI, SQLAlchemy, Jinja, Flask), SQL (mainly PostgreSQL, SQLite), git (GitLab & GitHub, incl. CI/CD pipelines), Vue.js, bash, AWS (incl. S3, SQS, CloudWatch), Argo Workflows, Sentry, Terraform, Docker, Kubernetes

Data Science

A "data detective" who enjoys learning about new domains, particularly natural sciences and engineering, by diving deep into data and sharing insights through effective visualizations.

- PhD in Machine Learning; worked with both structured data (incl. time series) and unstructured data (text documents & images) and developed novel algorithms and neural network architectures
- designed and implemented data products, incl. gathering requirements, exploratory data analysis, building, evaluating, and interpreting predictive models, and creating insightful data visualizations, e.g., to analyze chemical process data at BASF and alcemy

Tools: Python (incl. Jupyter, streamlit, pandas, sklearn, pytorch, numpy, scipy, plotly, matplolib, seaborn, dash), Grafana, MatLab, R

Communication

Proficient in professional writing and public speaking, enthusiastic about structuring information, clarifying complex concepts with diagrams, and meticulous editing.

- written a free online book "A Practitioner's Guide to Machine Learning"
- taught 45+ machine learning courses to various audiences, ranging from department heads to aspiring data scientists

Tools: LaTeX, AsciiDoc(tor), HTML/CSS, Affinity Designer, office applications
Languages: German (native), English (fluent), French (basics), Spanish (basics)

EDUCATION

Ph.D. (Dr. rer. nat.) Computer Science | ML Group, TU (Technical University) Berlin — 04/2015 - 04/2020

- in the machine learning group of Prof. Dr. Klaus-Robert Müller; funded by the Elsa Neumann scholarship from the universities of Berlin
- thesis: Similarity Encoder - A Neural Network Architecture for Learning Similarity Preserving Embeddings: developed a novel NN architecture to map high dimensional data into a low dimensional embedding space, where arbitrary pairwise relations between the data points are preserved as the embedding vectors factorize a given target similarity matrix
- supervised bachelor and master students
- graduated magna cum laude

M.Sc. Computer Science | TU Berlin — 10/2012 - 03/2015

- focus: intelligent systems (machine learning, big data) & computational neuroscience (at the BCCN)
- thesis: Knowledge Extraction from Complex Biological Texts: A Machine Learning Approach
- graduated top of my class (1.0 on a scale from 1 (best) to 5)

B.Sc. Cognitive Science | University Osnabrück — 10/2009 - 09/2012

- interdisciplinary study program including courses in neurobiology, computer science, psychology, artificial intelligence, mathematics, computational linguistics, neuroinformatics, and philosophy; taught in English
- thesis: Comparing and Combining Multiple EEG Features in Motor Imagery BCI - A Large Scale Study
- graduated with distinction (1.1 on a scale from 1 (best) to 5)

Abitur (secondary school) | Fichte-Gymnasium Karlsruhe — 09/2000 - 06/2009

- 11th grade as a year abroad in Missouri (USA)

PUBLICATIONS

Exploring Word Usage Change with Continuously Evolving Embeddings
Franziska Horn
In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: System Demonstrations*, pages 290–297, Online, August 2021. Association for Computational Linguistics (ACL).

Forecasting Industrial Aging Processes with Machine Learning Methods
Mihail Bogojeski, Simeon Sauer, Franziska Horn, Klaus-Robert Müller
Computers and Chemical Engineering, 144:107123, 2021.

The autofeat Python Library for Automatic Feature Engineering and Selection
Franziska Horn, Robert Pack, Michael Rieger
ECML PKDD Workshops 2019, Springer, Cham, 2020.

Automating the search for a patent's prior art with a full text similarity search
Lea Helmers*, Franziska Horn*, Franziska Biegler, Tim Oppermann, Klaus-Robert Müller
PLoS ONE, 14(3):e0212103, 2019.

Predicting Pairwise Relations with Neural Similarity Encoders
Franziska Horn, Klaus-Robert Müller
Bulletin of the Polish Academy of Sciences: Technical Sciences, 66(6):821-830, 2018.

Context encoders as a simple but powerful extension of word2vec
Franziska Horn
In *Proceedings of the 2nd Workshop on Representation Learning for NLP*, pages 10–14, Vancouver, Canada, August 2017. ACL.

"What is Relevant in a Text Document?": An Interpretable Machine Learning Approach
Leila Arras, Franziska Horn, Gregoire Montavon, Klaus-Robert Müller, Wojciech Samek
PLoS ONE, 12(8):e0181142, 2017.

Explaining Predictions of Non-Linear Classifiers in NLP
Leila Arras, Franziska Horn, Gregoire Montavon, Klaus-Robert Müller, Wojciech Samek
In *Proceedings of the 1st Workshop on Representation Learning for NLP*, pages 1–7, Berlin, Germany, August 2016. ACL.

Robust Artifactual Independent Component Classification for BCI Practitioners
I. Winkler, S. Brandl, F. Horn, E. Waldburger, C. Allefeld, M. Tangermann
Journal of Neural Engineering, 11(3):035013, 2014.

Predicting Treatment Response in Social Anxiety Disorder From Functional Magnetic Resonance Imaging
O. Doehrmann, S. S. Ghosh, F. E. Polli, G. O. Reynolds, F. Horn, A. Keshavan, ... & J. D. Gabrieli
JAMA Psychiatry, 70(1):87-97, 2013.

Increasing the Spectral Signal-To-Noise Ratio of Common Spatial Patterns
Franziska Horn, Sven Dähne
Proceedings of the Fifth International Brain-Computer Interface Meeting, 2013.

Combining Multiple EEG Features in Motor Imagery BCI
Franziska Horn, Johannes Höhne, Sven Dähne, Benjamin Blankertz
BBCI Workshop - Advances in Neurotechnology, Berlin, Germany, 2012.

PREPRINTS

The DALPHI annotation framework & how its pre-annotations can improve annotator efficiency
Robert Greinacher, Franziska Horn
arXiv preprint arXiv:1808.05558, 2018.

Discovering topics in text datasets by visualizing relevant words
Franziska Horn, Leila Arras, Gregoire Montavon, Klaus-Robert Müller, Wojciech Samek
arXiv preprint arXiv:1707.06100, 2017.

Exploring text datasets by visualizing relevant words
Franziska Horn, Leila Arras, Gregoire Montavon, Klaus-Robert Müller, Wojciech Samek
arXiv preprint arXiv:1707.05261, 2017.

Interactive Exploration and Discovery of Scientific Publications with PubVis
Franziska Horn
arXiv preprint arXiv:1706.08094, 2017.