Dr. Franziska Horn

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

A **data scientist** and **software engineer** turned **product manager** with 11+ years of experience
building data products in both research and application contexts.

SKILLS

Leadership

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

PRODUCT

• defined team vision, strategies, and goals for the data and solutions engineering team at alcemy
• identified user needs, gathered requirements, and designed solutions for new data products (e.g., internal analysis tools at alcemy and various customer projects as a freelancer)

PROCESSES

• optimized and designed new processes, incl. automation of many steps, e.g., to improve the customer onboarding at alcemy (reduced lead time from 50 to 25 days)
• conducted workshops to enhance knowledge sharing practices and encouraged more extensive documentation, resulting, for example, in an improved onboarding experience for new hires at alcemy
• implemented a new ticket board in Notion to streamline our team's sprint planning process at alcemy

PEOPLE

• mentored junior team members and students and conducted 1:1 data science coaching sessions for domain experts as a freelancer
• capacity planning and hiring of new team members (incl. writing job descriptions, devising take home tasks, conducting technical and culture fit interviews, onboarding), building up the ML team at Spectrm and the solutions engineering team at alcemy
• gave feedback and conducted performance reviews as the head of solutions engineering at alcemy

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; developed novel algorithms and neural network architectures and worked with both structured data (incl. time series) and unstructured data (text documents & images)
• designed and implemented data products, incl. exploratory data analysis, building and evaluating 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

Software Engineering

Passion for clean code, crafting software architectures with optimal complexity, and refactoring existing code as needed.

• designed and implemented various web apps from scratch (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 400+ stars on GitHub
• planned and executed a series of targeted refactorings to minimize code required to implement the data ingest for new customers at alcemy
• worked on a shared codebase at alcemy, incl. contributing to production code and reviewing merge requests

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

Communication

Proficient in professional writing and public speaking, enthusiastic about structuring information, clarifying 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
• authored several scientific papers published in journals and presented at conferences
• gave a talk at the PyData conference 2019 in Berlin in front of over 100 people

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

WORK EXPERIENCE

Freelance Data Product Consultant │ Leipzig (Germany) / remote   10/2018 - present
- 1:1 coaching sessions and multi-day workshops on how to use machine learning techniques in practice
- design, implementation, and evaluation of data science solutions tailored to my client's needs
- web development with a focus on data-powered web apps (using Python, incl. FastAPI & streamlit)
- clients include BASF and TRUMPF

Head of Solutions Engineering │ alcemy, Berlin   10/2022 - 11/2023
- fighting climate change, one cement and concrete plant at a time!
- hired and led a team of three data scientists and engineers, responsible for customer integration and resolving data-related customer requests
- created team vision & strategies and prioritized projects & objectives together with the customer success, product, and sales teams
- streamlined internal processes, e.g., halved onboarding time for new cement customers, minimized required code files per customer to 6 from 17
- built data products empowering the customer success team to conduct recurring customer data analyses independently

Senior Customer-Facing Data Scientist │ alcemy, Berlin   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 integration of new customers into our cement product by planning and implementing refactorings of core product components

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

Data Science Consultant │ BASF, Ludwigshafen   03/2018 - 10/2018
- most experienced Python developer in the team, responsible for code review
- 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
- in the machine learning group of Prof. Dr. Klaus-Robert Müller
- 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

Machine Learning Scientist │ Spectrm, Berlin   07/2016 - 06/2017
- developed a chatbot "AI" to respond to user messages automatically (using RiveScript)
- implemented a content recommendation API for newspaper articles, which can be used by all clients to promote their content (Python Flask App)
- established the machine learning team, incl. hiring two ML engineers

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 algorithms for curated shopping)

Student Research Assistant │ ML Group, TU Berlin   09/2013 - 09/2014
- focus on text classification, unsupervised learning (word2vec vector space 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 with results 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

EDUCATION

Ph.D. (Dr. rer. nat.) Computer Science │ TU (Technical University) Berlin · · · · · · · · · · 04/2015 - 04/2020

- in the machine learning group of Prof. Dr. Klaus-Robert Müller
- 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; potential application areas include recommender systems, especially to target the cold start problem, i.e., generate recommendations for items without previous user ratings, and natural language processing, by extending the word2vec algorithm to produce embeddings for out-of-vocabulary words and words with multiple meanings
- supervised bachelor and master students
- funded by the Elsa Neumann scholarship from the universities of Berlin
- graduated magna cum laude

M.Sc. Computer Science │ TU Berlin · 04/2013 - 03/2015

- focus: intelligent systems (machine learning, big data & Hadoop, neurobiology)
- thesis: Knowledge Extraction from Complex Biological Texts: A Machine Learning Approach (supervisor: Prof. Dr. Klaus-Robert Müller, TU Berlin)
- graduated top of my class (1.0 on a scale from 1 (best) to 5)

(M.Sc. Computational Neuroscience) │ BCCN / TU Berlin · · · · · · · · · · · · · · · · · · 10/2012 - 03/2013

- interdisciplinary & strongly research oriented international master program with a highly competitive application process (10 places/year)
- switched to computer science after 1 semester to deepen my technical knowledge and get a wider choice of application areas

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 in the field of brain-computer interfaces at the TU Berlin: Comparing and Combining Multiple EEG Features in Motor Imagery BCI - A Large Scale Study (supervisor: Prof. Dr. Benjamin Blankertz, TU Berlin)
- 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.

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. Association for Computational Linguistics.

"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. Association for Computational Linguistics.

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.