

Dr. Martijn Arts

Freelance Data Engineer · Python Software Engineer · Data Scientist

freelance@martijnarts.de | [LinkedIn](#) | [GitHub](#)

Freelancer for **Data Engineering**, **Python Software Engineering** and **Data Science**, backed by a strong machine learning and statistics background. Industry experience in consulting and in leading data projects through the entire product life cycle — from ideation and proof-of-concept to bringing data products into production in the cloud and operating them.

CORE SKILLS

Languages & core	Python (Pandas, Polars, Numpy, Pydantic, FastAPI), SQL, Bash
Data & cloud	Snowflake, dbt, Postgres, Spark / Databricks, AWS (S3, Batch, Lambda, Step Functions, EventBridge Scheduler, CloudFormation), Azure (AzureML, DevOps Pipelines, Application Insights), Docker, Kubernetes, Datadog
Machine learning	Scikit-learn, XGBoost, Prophet, BoTorch, Scikit-Optimize, MLflow — statistical modelling, time-series forecasting, black-box optimization, model calibration & evaluation
AI-assisted development	GitHub Copilot, Claude Code, Anthropic (Claude) and OpenAI models — code migration, refactoring, test generation and day-to-day development
Engineering practice	Unit & integration testing (Pytest), CI / CD, Git, MLOps, monitoring & alerting, code review, refactoring and stabilization of large code bases
Ways of working	Agile team lead (SCRUM), requirement & stakeholder management, project management, mentoring, presenting to executive and non-technical audiences

PROFESSIONAL EXPERIENCE

Jan 2022 — now	Freelancer — Data Engineering, Python Software Engineering & Data Science Independent consultant for clients in energy trading and central banking — see <i>Selected Projects</i> below.
Aug 2018 — Dec 2021	Senior Data Scientist , Henkel AG & Co. KGaA, Düsseldorf, Germany <i>In-house data science consulting covering a broad range of use-cases across the organization</i> - Led and contributed to data science / machine learning POCs and products in diverse business contexts: customer credit risk assessment, smart debt collection, product quality optimization, chemical experiment design and formulation optimization, energy demand forecasting and hair health assessment from sensor measurements. - Mentored junior data scientists on machine learning techniques, infrastructure, deployment and scaling; consulted use-cases in the ideation phase and conducted quality assurance on projects run with external implementation partners. - Led the initiative to standardize the team's data science toolchain: requirement analysis, evaluation of AzureML and Databricks, deployment of an experimental AzureML environment, demo application and team training sessions. - Advocated software engineering and DevOps principles in machine learning projects and pioneered Spark (Databricks) CI / CD workflows including unit and integration testing, in close collaboration with Henkel's DevOps team. - Contributed to digital upskilling in the Finance department with two presentations on machine learning use-cases broadcast to hundreds of employees; created the team's data challenge for recruiting and conducted technical interviews. Technologies: Python, Pandas, Numpy, Scikit-learn, XGBoost, Spark, Databricks, Prophet, MLflow, AzureML, Matplotlib, Seaborn, Plotly Dash, BoTorch, Scikit-Optimize, MS SQL, Docker, Kubernetes, Brigade, Azure, Azure DevOps Pipelines, Git, Linux
Apr 2017 — Jul 2018	Data Scientist , Qlaym GmbH, Düsseldorf, Germany <i>Data science consulting and machine learning backend development in a data science startup</i> - Exploratory data analysis of customer data such as financial time-series and sensor data from chemical production plants; built time-series forecasting models to predict sales for a customer in the agricultural sector. - Developed a new anomaly detection sub-module for the in-house data science platform and worked agile (SCRUM) on its machine learning backend. - Set up and maintained continuous integration for the platform with Jenkins and Docker; contributed to the Dask open-source project by reporting and fixing bugs. Technologies: Python, Pandas, Numpy, Scikit-learn, XGBoost, Prophet, Matplotlib, Seaborn, Dask, Celery, Docker, Jenkins, MongoDB, SQL (Exasol), Redis, RabbitMQ, Lua, Kafka, Git, Linux, SCRUM

Dec 2011

–

Mar 2017

Scientific Staff, RWTH Aachen University, Aachen, Germany

Research on detection algorithms for wireless communications at the Institute for Theoretical Information Technology

- Participated in 2 DFG-funded research projects, contributed to 3 research grant applications and published at international conferences and journals; received the best paper award at ICUFN 2016, Vienna.
- Supervised master and bachelor theses as well as student research assistants; developed proficiency in public speaking and teaching over 8 semesters of university teaching and 3 international conferences.
- Gained profound knowledge of the analysis of large datasets and distributed Monte Carlo simulations, statistical hypothesis testing, detection and estimation theory, signal processing, information theory, stochastics, convex optimization and compressed sensing.

Technologies: MATLAB, Python, C, LaTeX, Git, Linux, MacOS

2007

–

2011

Earlier positions — RWTH Aachen University & Philips Research, Eindhoven, Netherlands

Graduate and student research assistant at RWTH Aachen: hardware development for a fully digital magnetic induction measurement system (Chair for Medical Information Technology) and software development for a railway capacity allocation simulator. Internship at Philips Electronics Netherlands B.V. (Biomedical Sensor Systems): neonatal monitoring research covering hardware specification and development and a C++ framework for multi-sensor data acquisition and analysis.

Technologies: C / C++, Python, VHDL, Cern ROOT, Programmable Logic, Electronics

SELECTED PROJECTS

Mar 2026

–

Sep 2026

Orchestration Software for a Fundamental Energy Market Simulation Model

Industry: Energy Trading — European energy trading company **Role:** Python Software Engineer / Data Engineer

In-house software that orchestrates a third-party fundamental model simulating the energy market. It ingests time-series from 13 data providers, applies user-configurable processing and calculations, uploads the prepared inputs to the simulation software, supervises the simulation runs and downloads and post-processes the results. Scheduling of runs, data processing and calculations are extensively configurable through a UI. The software produces 14 simulation runs per day across 9 scenarios and more than 20 sensitivities covering short to medium-term time horizons; results are published within the company data platform for use in other projects and dashboards.

- Reduced production incidents from daily to 1–2 per week within less than a month by improving transparency through logging and reducing technical debt
- Took over the project and carried out refactoring and stabilization of the existing code base
- Migration of a large, monolithic C++ component of the software to Python, accelerated by LLM-assisted development (GitHub Copilot, Anthropic and OpenAI models)
- Bug fixing and implementation of new feature requests in close collaboration with analysts
- Maintenance and supervision of the scheduled simulation runs, post-processing and publication of results to the data platform

Technologies: Python, Pandas, Pydantic, Pytest, Postgres, AWS (S3, Batch, Lambda, Step Functions, EventBridge Scheduler), Serverless, CloudFormation, Datadog, Azure DevOps

Jun 2024

–

Sep 2026

Design and Implementation of the Backend for a Power & Commodity Market Report

Industry: Energy Trading — European energy trading company **Role:** Data Engineer / Python Software Engineer

The Power and Commodity Market Report is a tool for analysts and traders to analyze forecasts from various providers alongside current market data to support trading activities and investment decisions. It superseded two manually maintained reports, replacing several days of manual work per quarterly update with dynamic updates produced within hours whenever new data becomes available, and was re-designed from scratch.

- Design and implementation of the report backend based on Python, Snowflake and AWS, later supported by LLM-assisted development (GitHub Copilot, Anthropic and OpenAI models)
- Ingestion of long and medium-term forecasts and power market data from 10 external providers and internal sources via files, APIs and databases
- Harmonization, processing, aggregation and mapping of data from various providers to achieve comparability between their forecasts and market data, including currency conversion, inflation adjustments and mapping of technical terms and geographic definitions
- Definition and implementation of the data model for Snowflake using dbt, and automation of the ingestion and harmonization jobs
- Close collaboration with subject-matter experts (analysts) and front-end developers

Technologies: Python, Polars, Pandas, Pydantic, FastAPI, Snowflake, dbt, SQL, Pytest, AWS (Batch, Lambda, Step Functions), Serverless, CloudFormation, Datadog, Azure DevOps

Jan 2022

–

May 2024

Implementation of a Data Science Platform / Migration of Data & Business Processes

Industry: Finance and Insurance / Monetary Authorities — European Central Bank (ECB) **Role:** Team Lead (from June 2023) / Data Engineer / Python Software Engineer

Large-scale migration project for the ECB: development of a new data platform holding tens of millions of time-series, and migration of all data & business processes from the legacy system. Several agile development teams worked together on a large Python code base (>100,000 lines of code).

- Team lead of an agile team of 10 developers practicing SCRUM (previously Senior Developer)
- Definition, implementation and maintenance of a calculation & validation engine for time-series data including data processing & aggregation, statistical estimations and forecasting
- Development and maintenance of a common function library
- Migration of more than 50 data & business processes from the legacy system to the new platform
- Development of unit tests for all components (engine, library and data / business processes)

Technologies: Python, Pandas, Numpy, Pytest, Unittest, Kubernetes, NetworkX, Numba, Camunda, Jira, Confluence, GitLab, SCRUM

Sep 2021

–

Dec 2021

Design of Chemical Experiments and Formulation Optimization via Black-Box Optimization

Company: Henkel AG & Co. KGaA — FMCG **Role:** Data Scientist

Support of product developers in R&D with design of experiments and optimization of chemical formulations using a custom developed software utilizing black-box optimization.

- Contributed to design and implementation of a benchmarking framework to test algorithm candidates
- Conducted benchmarks on algorithms, documented findings and presented feedback to the product team
- Contributed to code clean-up and a major refactoring of the codebase, and wrote unit tests to increase test coverage

Technologies: Python, Pandas, Numpy, Pytest, BoTorch, Scikit-Optimize, AzureML

Jan 2021

–

Jul 2021

Data Science Toolchain / Platform Standardization

Company: Henkel AG & Co. KGaA — FMCG **Role:** Data Scientist, Machine Learning Engineer

Initiative to standardize the data science toolchain / platform within the team and to provide requirements and feedback for improving Henkel's data ecosystem in the long term.

- Researched machine learning operations approaches and interviewed data scientists across Henkel to gather requirements; exchanged ideas and aligned options with the data platform team
- Conducted an evaluation of Databricks and AzureML against the collected requirements
- Deployed and configured an experimental AzureML workspace and contributed to the professionalization and automation of AzureML workspace deployments
- Implemented a demo application covering common machine learning tasks such as data pre-processing and model training including experiment tracking and versioned model storage, and conducted training sessions for the team

Technologies: AzureML, Databricks, Python, Pandas, Scikit-Learn, Spark, MLflow

Jan 2019

–

Mar 2021

Predictive Credit Risk Assessment — from proof-of-concept to productionized data product

Company: Henkel AG & Co. KGaA — FMCG **Role:** Data Scientist, Data Engineer, Machine Learning Engineer, Project Manager

Credit risk assessment of customers based on a machine learning model that predicts the probability of default from internally available data, covering a receivables portfolio of several million open items. The model complements external rating information for bad-debt provisioning according to IFRS 9. Started as a proof of concept in 2019, taken into production as a fully operated data product and maintained until the end of 2021.

- Project management (budget, time booking, security, GDPR and architecture processes) as well as requirement and stakeholder management
- Definition of required data sources and provisioning by the BI team, including the necessary change requests
- Data exploration, data modelling and feature engineering of transactional data
- Design and implementation of the machine learning model on unbalanced classes including probability calibration, detailed performance analysis and consultation on expected impact
- Implementation of the feature extraction pipeline including unit and integration tests in Spark and pytest on Databricks
- Design and implementation of the machine learning operations architecture on Azure in close collaboration with the DevOps team: CI / CD pipelines, scheduling and orchestration of the machine learning pipeline with Brigade, automated model training on Kubernetes, performance tracking & model versioning in AzureML, logging & alerting with Azure Application Insights and Azure Monitor
- Communication of proof-of-concept results to key stakeholders and presentation of the concept to external auditors including technical correspondence
- Hypercare support after go-live; presented two live-streamed upskilling sessions on machine learning & AI based on this project to an audience of hundreds of internal employees

Technologies: Python, Pandas, Numpy, Spark, Scikit-Learn, XGBoost, Pytest, Matplotlib, Tox, Sphinx, Databricks, AzureML, MLflow, Brigade, Docker, Kubernetes, Azure DevOps Pipelines, Azure Application Insights, Azure Monitor

Jul 2019

–

Sep 2019

Energy Demand Forecasting

Company: Henkel AG & Co. KGaA — FMCG **Role:** Data Scientist, Data Engineer

Energy demand forecasting for optimizing energy procurement: future energy demand of a production line was forecasted using predictive models based on the production schedule and sensor measurements.

- Requirement and stakeholder management, data exploration and feature engineering
- Trained time-series machine learning models (scikit-learn, XGBoost, Prophet) and evaluated their performance
- Created an interactive dashboard in Plotly Dash presenting insights from the data exploration and visualizing model performance

Technologies: Python, Pandas, Scikit-Learn, XGBoost, Prophet, Matplotlib, Plotly Dash

Sep 2018

–

Dec 2018

Smart Accounts Receivables

Company: Henkel AG & Co. KGaA — FMCG **Role:** Data Scientist, Machine Learning Engineer

Optimization of collection efforts by assigning dunning strategies on customer level using a trained classifier.

- Data extraction from an MS SQL source database, data exploration and data modelling; feature engineering to capture customer payment behavior
- Training of a machine learning model on unbalanced classes, performance evaluation and presentation of proof-of-concept results to key stakeholders
- Mentored colleagues during implementation and roll-out of the product in 2019 / 2020

Technologies: Python, Pandas, Numpy, Scikit-Learn, Imblearn, Matplotlib, SQL

Aug 2018

–

Mar 2019

SalonLab

Company: Henkel AG & Co. KGaA — FMCG **Role:** Data Scientist, Machine Learning Engineer

The SalonLab Smart Analyzer is a handheld device that scans hair structure through near-infrared measurements. Hairdressers use the corresponding SalonLab App in their consultation process, where machine learning models assess hair health from the infrared measurements.

- Collaborated in a high-impact project with multiple external partners and internal teams and consulted on the design and development of the machine learning models
- Stress tested a commercial deployment solution for the model prediction microservice
- Collaborated with a cloud engineer to improve the deployment architecture, reducing prediction latencies from seconds to milliseconds and saving six figures in license costs compared to the commercial solution
- Implementation of the model prediction microservice exposing a REST API in Python including logging and error reporting; performed upgrades of the API in Dec 2019, Feb 2020 and Jul 2020

Technologies: Python, Flask, Gunicorn, Pandas, Docker, Kubernetes, Jira, Confluence, Apache JMeter

EDUCATION

Dec 2011

–

May 2017

Dr.-Ing. Electrical Engineering and Information Technology (equivalent to PhD)

RWTH Aachen University, Aachen, Germany

Title: “Eigenvalue-Based Spectrum Sensing for Cognitive Radio: Change Detection Problems and Fundamental Performance Limits”

Focus: Statistical hypothesis testing for signal detection in wireless communications

Publications: [Google Scholar profile](#)

Oct 2005

–

Jun 2011

Diplom-Ingenieur Computer Engineering (equivalent to Master of Science)

RWTH Aachen University, Aachen, Germany

Title: “Development of a fully digital FPGA based magnetic induction measurement system”

Focus: Signal processing, wireless communications and medical technology — graduated in the top 20 % of Computer Engineering graduates of the academic year

PERSONAL DETAILS

Citizenship: German

Languages: German (native), English (fluent)