GEELEN.IO

Version 1.7

One-Pager

Roles
Data Scientist
Data Engineer
DevOps/SRE developer
Technical and Unit Leader
Product Owner/Manager

Expertise
AI/ML/LLM
Software Development
Containerization
HPC and HAC computing
Data-Lake, Data Warehouse and Data Mesh patterns
IoT/Embedded
Observability
CI/CD
OPC-UA, CAN, Profibus/net

Programming Languages
Python
SQL
Git
R (Statistics)
JavaScript
C-embedded/Platform.io

Frameworks/Tools/Technologies
Kubernetes/Docker/KVM
Tofu/Terraform
Ansible
Kafka, Pub/Sub, RabbitMQ
Postgres/Supabase
ArgoCD
ELK and Grafana Stack
Linux

Cloud/Compute
Cloud Native/OpenStack
GCP/AWS/Azure
Edge processing and NPUs
GPU accelerated workloads

Certifications
GCP Authorized Trainer
GCP Professional Data Engineer
CKAD, CKA, KCNA, KCSA
ArgoCD: Fundamentals
ArgoCD: GitOps at Scale
Terraform Associate (003)

For all 140+ certificates, please check LinkedIn

Pieter Geelen
pieter@geelen.io
https://geelen.io

Projects

Jan 2025 - March 2025	Rapid Prototyping 3D Printing
Dec 2024 - April 2025	AI System Engineering
Oct 2023—Sep 2024	Senior DevOps Engineer
Aug 2023–Oct 2023	AI Inventory Management App
Jun 2023–Aug 2023	Fashion and AI
May 2023–Jul 2023	Industry and Digital Transformation

Jobs

Nov 2022 – Mar 2023	Tech Lead Platform
Jan 2022 – Oct 2022	Product Owner IoT
Mar 2020 – Oct 2021	Manager Data and Technology
June 2019 – Mar 2020	Data Scientist
Jan 2018 – Mar 2019	Doctorate
Aug 2017 – Dec 2017	Internship
Aug 2015 – Feb 2017	Entrepreneur and Owner

Education

Feb 2017–Dec 2017	MSc: Information Management & Business Intelligence
Sep 2008–Jun 2014	BSc: Health Science (Epidemiology)
Aug 1995–Jul 2008	European Baccalaureate

Languages

German	Native
English	Native
Dutch	Native

Professional Experience

Jan 2025 - March 2025
Rapid Prototyping with 3D Printing
A water-cutting machine's rubber cover kept breaking within minutes, causing sand and water damage. A Proof of Concept tested 3D-printed replacements using Nylon with Carbon Fiber, extending durability from 5 minutes to over 20 hours (240x improvement). Next steps involve setting up in-house 3D printing capabilities, training, and production setup. Click here for more info.

Dec 2024 - April 2025
System Engineering for Generative AI workloads
A web content creator explored Generative AI and needed a production setup. The project included hardware and software implementation, covering networking, server setup, and Linux, Docker, CUDA, and NVIDIA Container Toolkit configuration. Multi-GPU utilization was tested, and automation, documentation, and integration testing were developed for ongoing support. Click here for more info.

Oct 2023 - Sep 2024
Interim Senior DevOps Engineer — Schäfer Shop
This project involves the largest Office Supplies reseller in Germany, that is rebuilding its cloud infrastructure using OpenStack. State-of-the-art tools like Kubernetes, Helm, Ansible, Terraform and ArgoCD are used for deployment. The scope includes data related topics too, like an Apache Hadoop/Spark migration.

Aug 2023 - Oct 2023
Inventory Management App with AI — Wiethe CGI
Consultancy project to design, plan and support the development of an Android Native app using Flutter and Supabase (Postgres). Core features included edge AI processing of imagery to QA business processes. Click here for more info.

Jun 2023 - Aug 2023
Fashion and AI — Wiethe CGI
Consultancy project to validate the current code-base of a newly developed application and support further development, architecture and documentation. Click here for more info.

May 2023 - Jul 2023
Industry and Digital Transformation — Nyrstar
Consultancy project for roadmapping digital transformation in the process industry focussing on OT and IT systems and what infrastructure can enable data-driven decision making within the organization. Click here for more info.

Nov 2022 - Mar 2023

Tech Lead Platform — Keelvar

As a former Tech Lead Platform at Keelvar, I held the responsibility of managing the platform team and supervising the development, maintenance, and enhancement of essential services for the sourcing and procurement SaaS platform. The scope included: DevOps engineering, Infrastructure as Code (IaC), and CI/CD processes.

Jan 2022 - Oct 2022

Product Owner, IoT — GrandCentrix

As a Product Owner in the IoT space, I was responsible for hardware and software development, manufacturing, production, and service delivery. Additionally, I was tasked with rescuing failing projects by connecting requirements, scope, and budget with development teams and stakeholders.

June 2019 - Oct 2021

Manager Data and Technology — Sitech

As a Data Scientist, I spearheaded the building of a data-science/-engineering team, achieving exceptional results that ultimately led to my promotion to Manager of Data and Technology. Notably, I developed an Auto-ML pipeline for the CBM solution, utilizing maintenance records and sensory data to deliver fast, cost-effective, and easily interpretable insights. Additionally, I designed and implemented a cloud-based event-driven solution in Azure specifically tailored for CBM. Other activities included R&D projects in the areas of IIoT and Smart Sensoring, focusing on edge-processing and hybrid modeling, while securing funding to support these initiatives.

Jan 2018 - Mar 2019

PhD Student — BISS

During my PhD candidacy, I conducted research on algorithmic transparency and GDPR-compliance at Business Intelligence & Smart Services (BISS) research organization. I presented my findings on an AI project at INFORMS in Phoenix, Arizona. Additionally, I supported projects that utilized computer vision and AI technologies by setting up the hardware and software for ML and AI.

Data-Science Portfolio

AutoML solution for Predictive Maintenance
Problem: Sitech Asset Health Center focussed on predictive maintenance, translating machine data into insights.

Solution: An automated ML pipeline was developed over time to create, deploy and maintain machine learning models. The different categories of models included:
For fouling, failure and remaining useful life:

- A Brute-Force method for finding the best features and algorithms to predict: fouling, failure and remainder useful life.
- A tree-search based modelling technique based on a knowledge graph to test the best algorithms with the best features to do the same.
- A genetic algorithm that had more freedom in the generation of features of algorithms and allowed for ensembles to generate too.

For detailing unknown events

- Unsupervised anomaly detection based on locality and isolation
- Unsupervised univariate likelihood methods like FFT KDE

For behavioural properties reduced data with the detection of

- Level Shift
- Persistence Shift
- Volatility Shift

And implemented those on a SQL level, so that these methods can be used with Grafana dynamically.

My role: was to underline the scientific basis for these methods, validate the results, lead the prototyping and supervise the production of this library. Special focus was put in the execution ability of the library. Massive parallel processing, in conjuncture with Kubernetes clusters and the deployment of these models into serverless technologies like Azure Functions, made the library realistic for its application.

Profitability Project Part 1
Problem: A chemical plant needed 4 times the amount of catalyser to produce the product than normal, rendering the production processes unprofitable. The dataset consisted of minute measurements over a year of machine data with a width of about 1000+ parameters.

Solution: The data was cleaned and further reduced to about 450 parameters. The use of libraries like Vaex was necessitated by these volumes. The behaviour was properly modelled with reinforcement methods like CatBoost, providing advantages of both GPU-accelaration and batch-wise learning, the latter being a requirement of experiment. CatBoost was also chosen due to its more generalized (naive) approach of developing a decision tree and great explainability of the model. With the integrated

SHAP library, we were able to identify the culprits that caused the problem by 1) evaluating feature importance and 2) reviewing the non-linear impact of certain parameters.

My Role: Scientific and Team Lead, prototyping was done by myself. The reporting was delegated.

Profitability Project Part 2

Problem: The first results were not "believed" by the engineers, and they asked us to develop a model that could not only outline the "offenders" of the process, but also recommend a solution. This required the use of causal inference that was able to use time series embedding, as changes in the system effect over time.

Solution: We used CausalNex library for this application because it supports both the structure learning of the graph, as well as determining the casualty of the vertices. We were able to consolidate the aforementioned "offenders" and found that there were autoregressive (oscillating) properties within the model. We were able to consolidate the findings with a chemical engineer and found there was a chemical saturation that was the root-cause of this problem.

My Role: Scientific and Team Lead. Allocated and supervised an internship position and recruited an Econometrics Intern to do a technical deep dive into the topic.

Hiccups of Steam Turbine

Problem: A steam turbine with very high RPM had the problem of hiccups. From hiccup to shut down the asset had about 1 to 5 seconds.

Solution: The project focussed on understanding whether the hiccups were measured with a current set of measurements that were situated on asset level. The measurements were done on millisecond interval and had a breadth of about 750 parameters. The data was cleaned and reduced. A spatial analysis showed that there were various instances of outliers, grouped in events. Extended Isolation Forest algorithm was used because of the ability to do massive parallel computation and batchwise learning of the ensemble, making it feasible for this used case. The distance was split into ideal groups using Otsu. The majority of the outlier groups coincided with the events of the hiccups and gave a 2-week lead time before failure. This meant that the turbine actually ran in a “failing mode” for 2 weeks before actually failing, and this space is represented in the data.

My Role: Scientific and Team Lead, prototyping was done by myself. The reporting was delegated.

Annotation Agreement in detecting Paul Ekmans in sound

Problem: BISS had an academic data-science project to detect emotion in sound. A decomposition of 5 seconds of sound was used to train an RNN to classify the emotions as defined by Paul Ekman. The snippets we also annotated by multiple individuals and aggregated through majority vote. The algorithm was struggling to converge, however, giving a poor Cronbach alpha as a result. When looking deeper into results, groups of disagreement could be observed.

Solution: I did a statistical analysis to prove, based on the annotations, that the emotions are not independent of one another, providing engineering challenges for training the algorithm, particularly on how to aggregate the annotations. This insight resolved the training issue of the algorithm and lead to a conference presentation at Institute for Operations Research and the Management Sciences or INFORMS in Phoenix, Arizona :

- Bromuri, S., Ebrahim, M., Iren, D., & Geelen, P. (2018). The Effect of Emotional Cues on Making Economic Decisions under Uncertainty. In INFORMS Annual Meeting INFORMS. http://www.abstractsonline.com/pp8/#!/4701/presentation/19115

My Role: As PhD candidate I was responsible for the design, execution and presentation of the findings.

Solar Panel Detection on Satellite Imagery

Problem: BISS had an academic data-science project to detect solar panels from satellite imagery. The AI model was somewhat generic, as it was yet another CNN with some fully connected layers at the back to learn the visual concepts and classify them accordingly. There was no setup of hardware/software to that was feasible at the time to do the project.

Solution: A server space was created, which included connectivity, power, racks and climate control. For the project, special attention went to:

- the safe storage of image data
- the preprocessing of vast quantities of data
- GPU-accelerated learning
- networking
- thermal management
- overall installation and configuration of Linux, Python and TensorFlow ecosystems.

My Role: As PhD candidate, it was my job to organize the annotation process and set up the training environment, including the requisitioning, installation and provisioning of servers.

When does cocoding occur?
Problem: BISS was doing research on the topic of cocoding. A GitLog was used to test for collaboration and blame.

Solution: The transactional data was transformed into an interaction matrix that was time-embedded. The hypothesis were that 1) that there is an underlying pattern (graph) of actors which causes interactions and 2) that this pattern is dependent on 2.1) Expertise of the actor and 2.2) Department in which the actor works. An empirical network toolbox was used to evaluate the results, called Sienna (https://www.stats.ox.ac.uk/~snijders/siena/). The massive computations required 700 to 1400 GB of Memory and posed serious challenges for the data-infrastructure.

My Role: As PhD candidate I was responsible for the design, execution and presentation of the findings.

Hyper Resolution for Epidemiologic Surveys
Problem: Epidemiologic Surveys are conducted at different granularities. A consultancy firm wanted to provide granular predictions on village, and even zip-code level, on the epidemiological status of a region.

Solution: The data was collected and evaluated. Based on the properties found and academic literature review that was conducted, a simple regression ensemble that focussed on Attributive Risk, was constructed as means for learning the patterns. The results were integrated into a BI dashboard and shared with customers.

My Role: Scientific Lead and Internship Support