



# Jannik Gut

- 30 June, 1996
- jannik.gut@gmx.ch
- 041 460 31 07
- Seesatz 39, 6204 Sempach
- [github.com/rostro36](https://github.com/rostro36)
- <https://rostro36.github.io/>
- [linkedin.com/in/jannik-gut](https://www.linkedin.com/in/jannik-gut)

## Languages

- English ●●●●●
- French ●●●●●
- German ●●●●●

## Programming

- Bash ●●●●●
- C ●●●●●
- Google Apps Script ●●●●●
- Python ●●●●●
- SQL ●●●●●

## Tools

- Azure ●●●●●
- Command Line ●●●●●
- Markdown ●●●●●
- git ●●●●●
- Google Suite ●●●●●
- Linux ●●●●●

## Hobbies

- Computers
- Cycling
- Godfather of a 9 year old
- Jogging
- Reading the newspaper

## Previous Career

- 2020 – 2021 **Master Computer Science** ETH Zürich  
Master thesis about disease representations in administrative claims data
- 2019 – 2020 **Data Analyst Intern** F. Hoffmann-La Roche  
Worked on the Competitive Intelligence project for Talent Acquisition, as part of the IT for HR department.
- 2015 – 2019 **Bachelor Computer Science** ETH Zürich  
Thesis about the security of the transmission of ballot voting results.
- 2011 – 2015 **Matura** Kantonsschule Sursee  
Matriculation project about the difference of news during the Crimea crisis between the Ukrainian side and the Russian side.

## Academical projects

- Advanced Machine Learning The lecture "Advanced Machine Learning" gives some tasks to solve, where my team always beat the hardest baseline. The first task were about data pre-processing, which we solved with sklearn, the second about class imbalance where imblearn was used and the last one about ECG data where BioSPPy and NeuroKit2 helped. Python
- Baum-Welch As a project in the course "Advanced Systems Lab" my team had to build a fast implementation of the Baum-Welch algorithm for solving Hidden Markov Models in C, benchmark it and we also provided valgrind output. C
- Computational Intelligence Lab The course "Computational Intelligence Lab" demanded a project for which my team chose the Twitter classification task, where we employed BERT and ALBERT and tested the effect of lexical normalization with MoNoise. BERT
- Big Data The lecture Big Data was a great lecture with many hands-on exercises for Hadoop, HDFS, MongoDB, Azure, Spark, MapReduce, XML, JSON and many more. I made many notes and saved them in this repo. Hadoop
- Master thesis My Master thesis in the Medical Data Science Group tried to find the influence of hierarchical information and co-occurrence information in administrative claim patient data. For this project euclidean and hyperbolic machine learning was used with PyTorch as well as graph neural networks with PyTorch geometric. PyTorch
- Machine Learning for Health Care For the homework of the "Machine Learning for Health Care" course I employed Tensorflow Keras in the first project about transferal learning of ECG data, different sklearn functions for the second NLP problem and again Tensorflow Keras in the last MRI segmentation problem. Python
- Partisan Responses For the GESS subject "Sequencing Legal DNA" my team used open information extraction after pre-processing with spaCy the U.S. congressional record to form small knowledge graphs and then use GraphWriter to generate responses to questions. As a benchmark GPT-2 answers were also used. Python



# Jannik Gut

- 30 June, 1996
- jannik.gut@gmx.ch
- 041 460 31 07
- Seesatz 39, 6204 Sempach
- [github.com/rostro36](https://github.com/rostro36)
- <https://rostro36.github.io/>
- [linkedin.com/in/jannik-gut](https://www.linkedin.com/in/jannik-gut)

## Languages

- English ●●●●●
- French ●●●●●
- German ●●●●●

## Programming

- Bash ●●●●●
- C ●●●●●
- Google Apps Script ●●●●●
- Python ●●●●●
- SQL ●●●●●

## Tools

- Azure ●●●●●
- Command Line ●●●●●
- Markdown ●●●●●
- git ●●●●●
- Google Suite ●●●●●
- Linux ●●●●●

## Hobbies

- Computers
- Cycling
- Godfather of a 9 year old
- Jogging
- Reading the newspaper

## Professional projects

Competitive Intelligence	This project uses <u>AWS EC2</u> and <u>RDS (MySQL)</u> instances to crawl and store talent acquisition data, afterwards use machine learning to categorize data and in the end use <u>Tableau</u> to present a dashboard. Some of the auxiliary functions have been automated with <u>Google Apps Script</u> (JavaScript-dialect). This project also required me to work with non-technical clients.	GAS
ERP Specification RPA	With the help of <u>Google Apps Script</u> I programmed a tool to create specifications for <u>Workday</u> integrations, which also simplified the automation of the impact assessment.	GAS
Splunk	The ERP software <u>Workday</u> needed a new monitoring solution, for which I helped integrating <u>Splunk</u> .	Splunk

## Leisure projects

### Books/Webinars

<u>Clean Code - Agile Software Craftsmanship</u>	Robert C. Martin
<u>Docker Mastery: The Complete Toolset</u>	Bret Fisher
<u>Python Testing with pytest</u>	Brian Okken
<u>Splunk Fundamentals 1</u>	Splunk

### GitHub

<u>Notes</u>	In this repo are notes on interesting NLP and knowledge graph papers as well as lecture notes of "Advanced Machine Learning" and "Big Data".	Markdown
<u>NLZMarkov</u>	Scrapes the search of the <u>Neue Luzerner Zeitung</u> for a search term and makes a <u>Markov chain</u> out them. There are options for part of speech tagging and a <u>Qt-Gui</u> .	Python
<u>TdFGraph</u>	Scrapes the official placements of <u>Tour de France</u> or <u>Vuelta a España</u> stages or from a fan page and displays the results with <u>matplotlib</u> .	Python
<u>TLSCheck</u>	Scrapes <u>Wikipedia</u> for Swiss municipality websites, checks the websites for <u>TLS</u> and displays the results with <u>matplotlib</u> .	Python

## References

Thesis Supervisor	Thomas Sutter (thomas.sutter@inf.ethz.ch)	PhD Student
Project Leader	Jingshu Sun (jingshu.sun@roche.com)	Data Scientist
Business Partner	Manuel Zumkeller El-Manchi (manuel.zumkeller@roche.com)	Business Partner