

02.01.25

Einführung in Data Science und maschinelles Lernen

ZEITREIHENANALYSEN

- **Fehlende Werte**
- **Muster in Zeitreihenanalysen**
- **Non-Stationarity**
- **Baseline Modelle und Naïve Forecasting**
- **Projektpräsentation**
- ***Und jetzt?***

EVALUATION VON IMPUTATIONEN

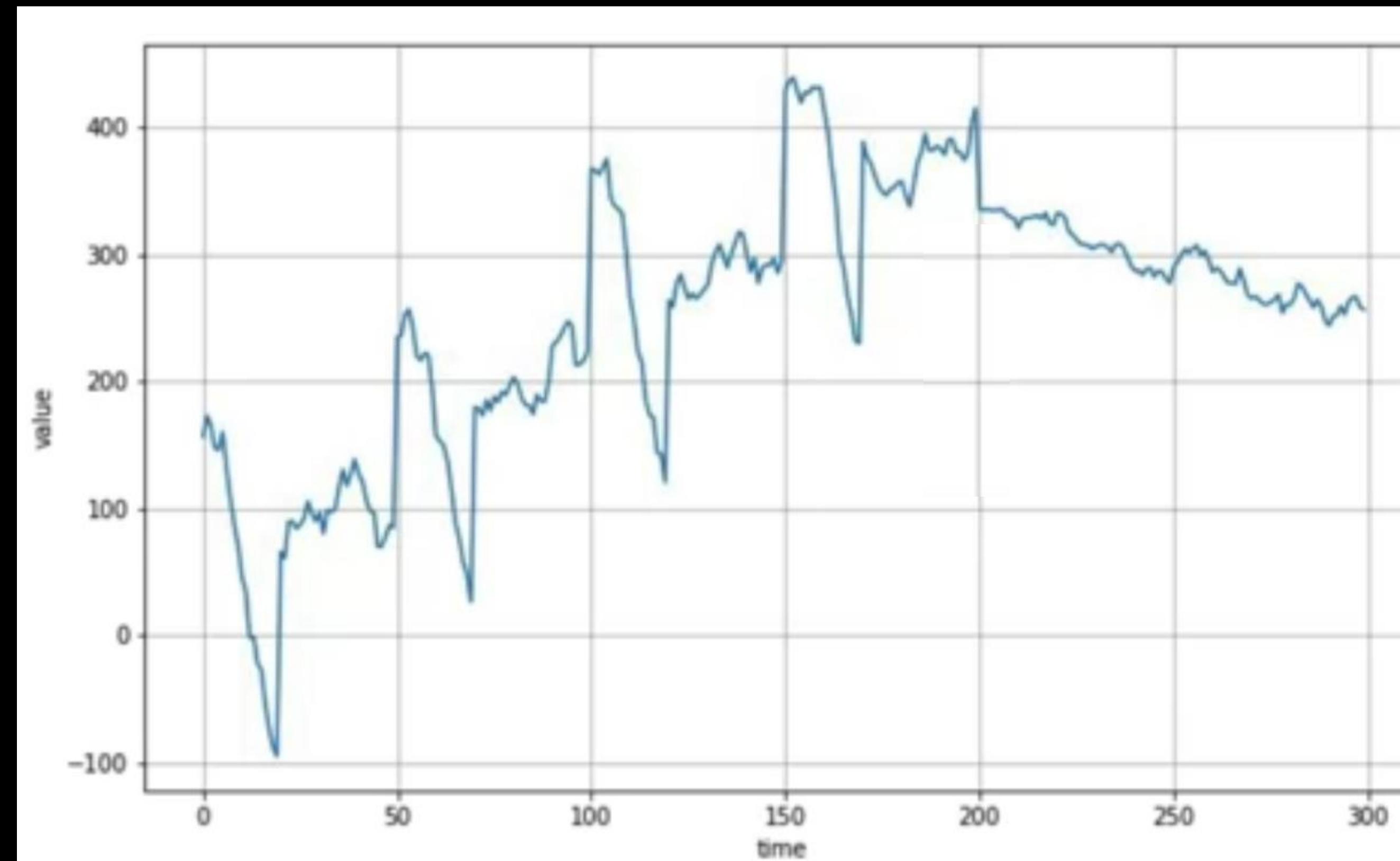
- **Betrachtung der Auswirkung auf die Vorhersage:
Berechnung der Vorhersagen für den Validierungsdatensatz
mit und ohne Imputationen**
- **Bewerten der Genauigkeit von imputierten Werten im
Vergleich zu bekannten Originalwerten:
Erstellen künstlicher Fehlwerte und Anwenden der
Imputationsmethode**

IMPUTATION BEI VIELEN FEHLENDEN WERTEN

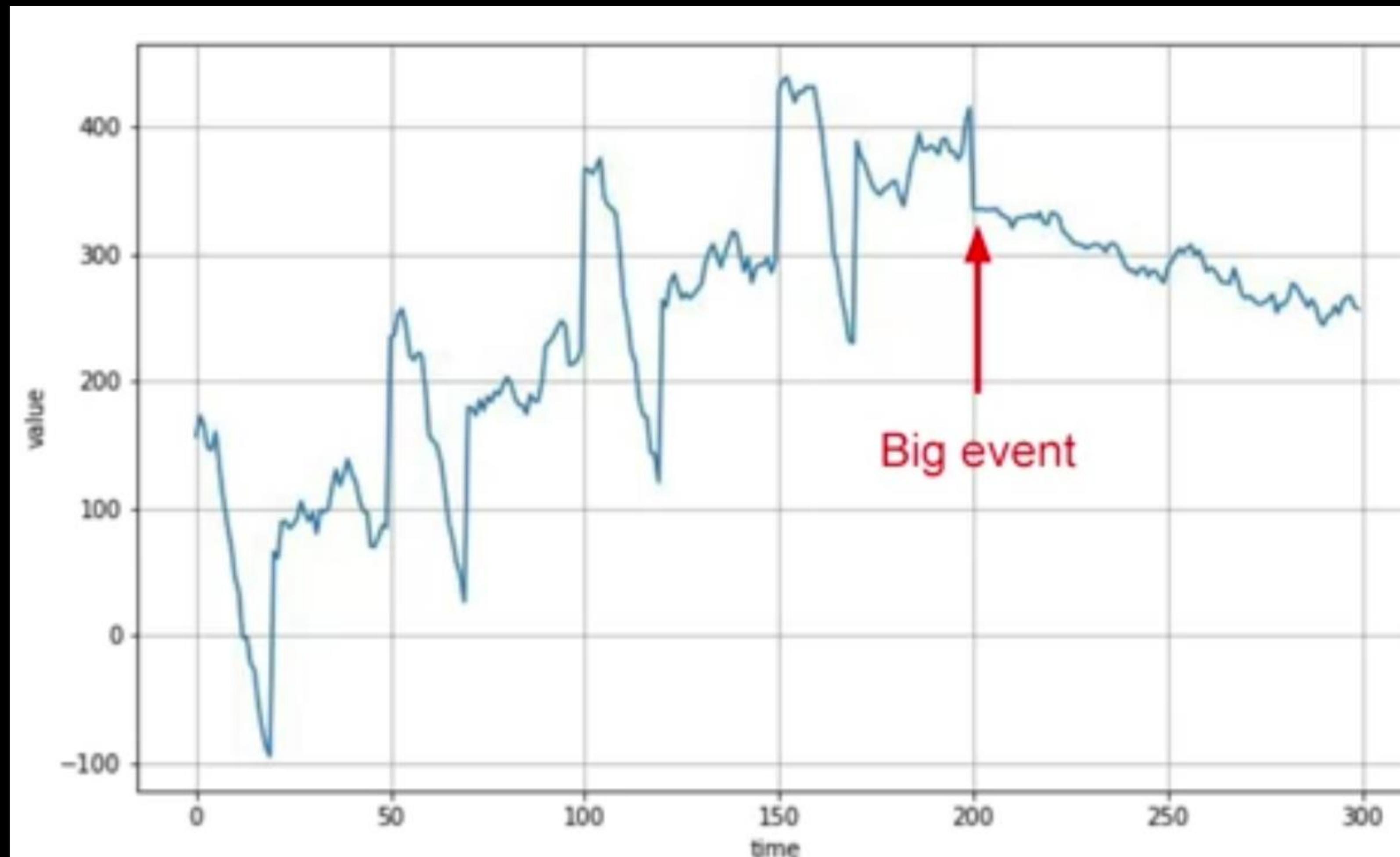
- Berücksichtigung der Information, welche Werte imputiert wurden, als Feature.

DISKUSSION

Welche verschiedenen Arten von Mustern kann man in der dargestellten Abbildung erkennen?



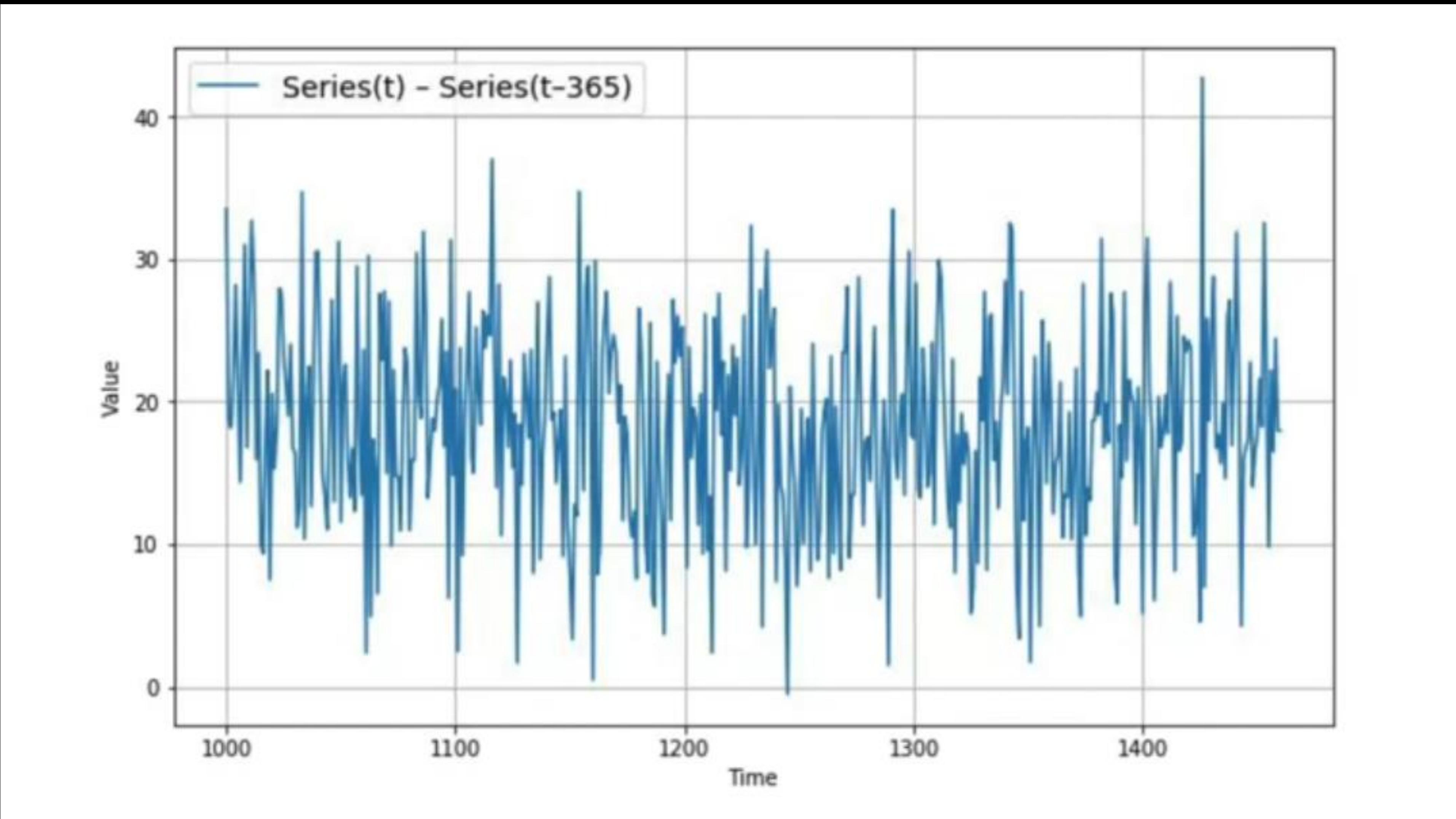
NON-STATIONARITY



MUSTER IN ZEITREIHENANALYSEN

- Trends
- „Jahresgang“ (Saisonicity)
- Rauschen (Noise)
- Autokorrelation (serielle Korrelation)

DIFFERENCING



DIFFERENCING

- **Subtraktion der vorherigen Beobachtung von der aktuellen Beobachtung**
- **Methode zum Entfernen von Trends und Saisonalität oder ggf. anderen Autokorrelationen in Zeitreihen**
- **Hilft, Autokorrelation zu reduzieren und stationäre Zeitreihen zu erhalten**

„LAG“

- Die Differenz zwischen aufeinanderfolgenden Beobachtungen wird als Lag-1-Differenz bezeichnet.
- Die Lag-Differenz wird an die spezifische zeitliche Struktur des Datensatzes angepasst.
- Bei Zeitreihen mit einer saisonalen Komponente wird der Lag üblicherweise auf die Dauer der Saisonalität gesetzt.

BASELINE MODELLE

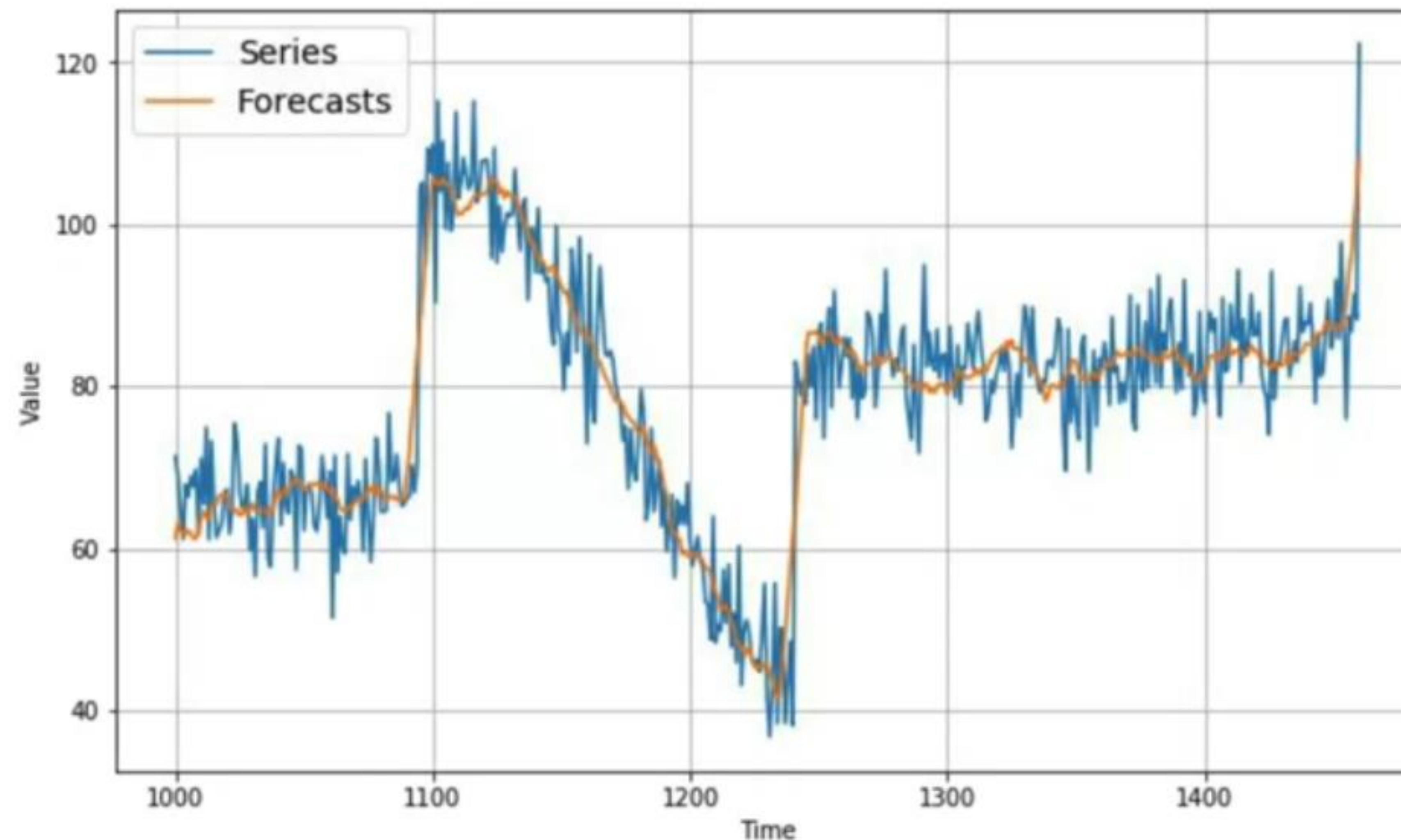
- Allein an Metriken wie MAPE oder RMSE kann man häufig schlecht abschätzen, wie gut das eigene Modell gelernt hat.
- Es ist daher wichtig, die Ergebnisse anderer Modelle als „Baseline“ bzw. Referenz zu nutzen.

Mögliche Baselines:

- Ergebnisse bisher genutzter Modelle für den gleichen Datensatz
- Ergebnisse von Modellen auf artverwandten Datensätzen
- Speziell bei Zeitreihen: Ergebnisse basierend auf Naïve Forecasting

NAÏVE FORECASTING

- Mögliches Baseline Modell für Zeitreihenanalysen
- Vorhersage entspricht dem jeweils letzten beobachteten Wert
- Saisonale Naïve Forecasting:
Vorhersage entspricht dem letzten Wert mit der gleichen Saisonalität.



Forecasts = trailing moving average of differenced series + centered moving average of past series ($t - 365$)

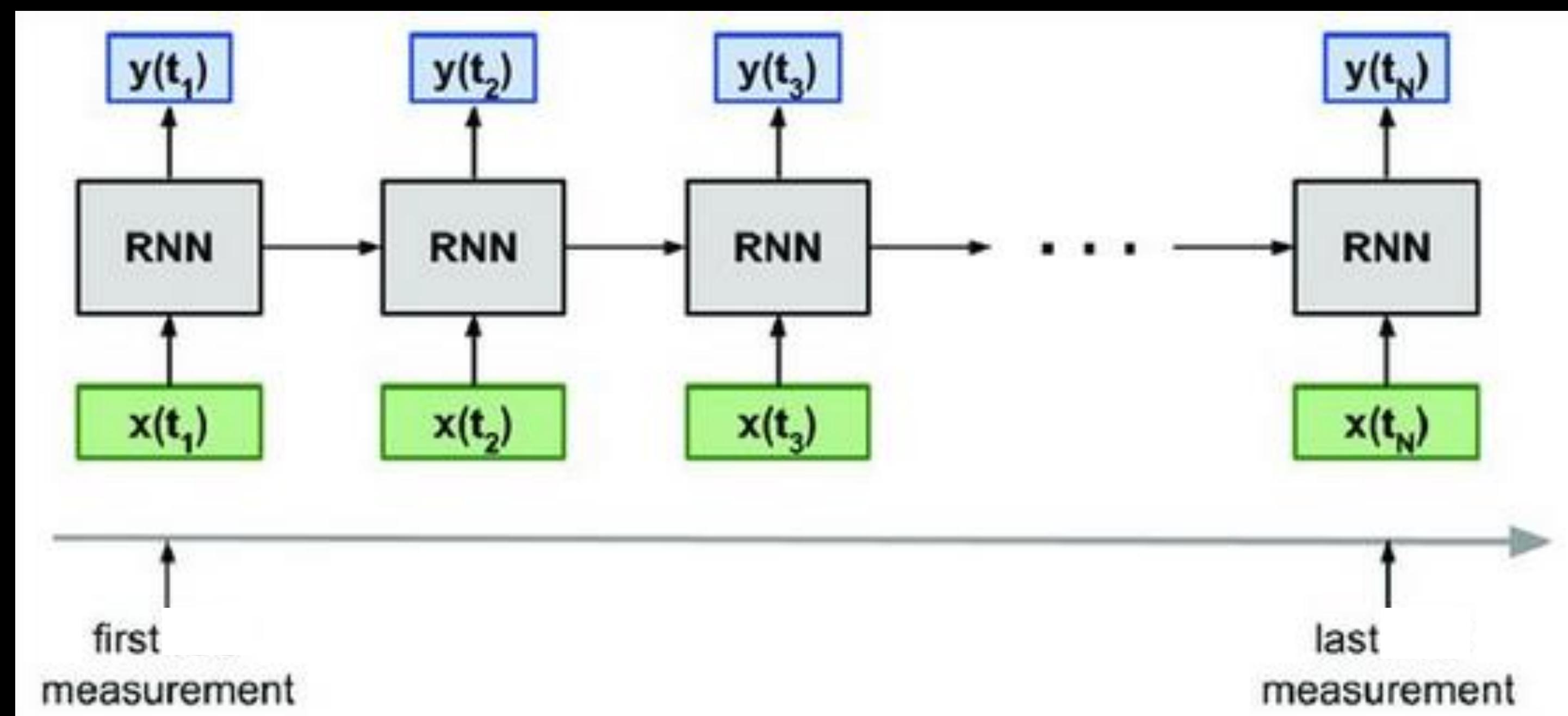
DISKUSSION

In der Praxis will man häufig nicht ein Jahr, sondern z.B. nur den nächsten Tag vorhersagen.

- Welche Informationen aus der Zeitreihe sind häufig sehr besonders hilfreich?

RECURRENT NEURAL NET (RNN)

- In einem RNN-Layer sind die Knoten einer Schicht untereinander verknüpft.



INHALT DER PROJEKTPRÄSENTATION

- **Eure Namen auf der Titelseite**
- **Auflistung und kurze Beschreibung der selbst erstellten Variablen**
- **Balkendiagramme mit Konfidenzintervallen für zwei selbst erstellte Variablen**
- **Optimierung des linearen Modells: Modellgleichung und adjusted r^2**
- **Art der Missing Value Imputation**
- **Optimierung des neuronalen Netzes:**
 - **Source Code zur Definition des neuronalen Netzes**
 - **Darstellung der Loss-Funktionen für Trainings- und Validierungsdatensatz**
 - **MAPEs für den Validierungsdatensatz insgesamt und für jede Warengruppe einzeln**
- **„Worst Fail“**

HINWEISE

- **Dauer der Präsentation: ca. 8 Minuten pro Team**
- **Powerpoint, Keynote oder ähnliches**
- **Bis 16 Uhr nächste Woche solltet Ihr für Euer bestes Modell die Vorhersagen für den Testdatensatz der Kaggle Competition berechnen und dort hochladen.**

Bis spätestens zum 1. März:

- **Die Präsentation zu Eurem Repo hinzufügen.**
- **Das Repo auf „public“ setzen.**
- **Das Repo wie in den READMEs beschrieben vervollständigen und das main README wie [hier](#) beschrieben in der EduHub-Plattform hochladen.**

WIE KANN ICH WEITER MACHEN?

=



KI-Forschung

27. Dezember 2024

Deepseek-V3: Ein 5,6-Millionen-Dollar-Wunder aus China mischt die KI-Elite auf



THE DECODER Newsletter

Die wichtigen KI-News direkt ins E-Mail-Postfach.

- ✓ 1x wöchentlich
- ✓ kostenlos
- ✓ jederzeit kündbar

E-Mail-Adresse *

Kostenlos abonnieren

Stay up to date in Data Science.

Get the Data Elixir newsletter for a weekly dose of the top data science picks from around the web. Covering machine learning, data visualization, analytics, and strategy.

Sign up for Free

and join more than **55,000**
data lovers today.

Email

SIGN UP

No spam, ever.

What readers say...



Caitlin Hudon
@beeonaposy



Replies to @beeonaposy and @joecrobak
Best wide-ranging sampler: Data Elixer

@lonriesberg compiles a broad look at what's happening in data, with titles like 'Modeling coronavirus. How to put R in production. Ray tips & tricks. Mathematics for ML. Finding your way in ML. Distrusting data.'



Julia Evans
@b0rk · Nov 11, 2018



do you subscribe to a mailing list you think is really useful? what is it?



Laurence Watson
@LaurenceWWatson

I find @dataelixir by @lonriesberg excellent; not too much and quality stuff



4



See Laurence Watson's other Tweets



Datenschutzerklärung ·
Nutzungsbedingungen

Explore topics

Machine Learning

Artificial Intelligence

Technology

Data Science

Deep Learning

AI

Programming

Python

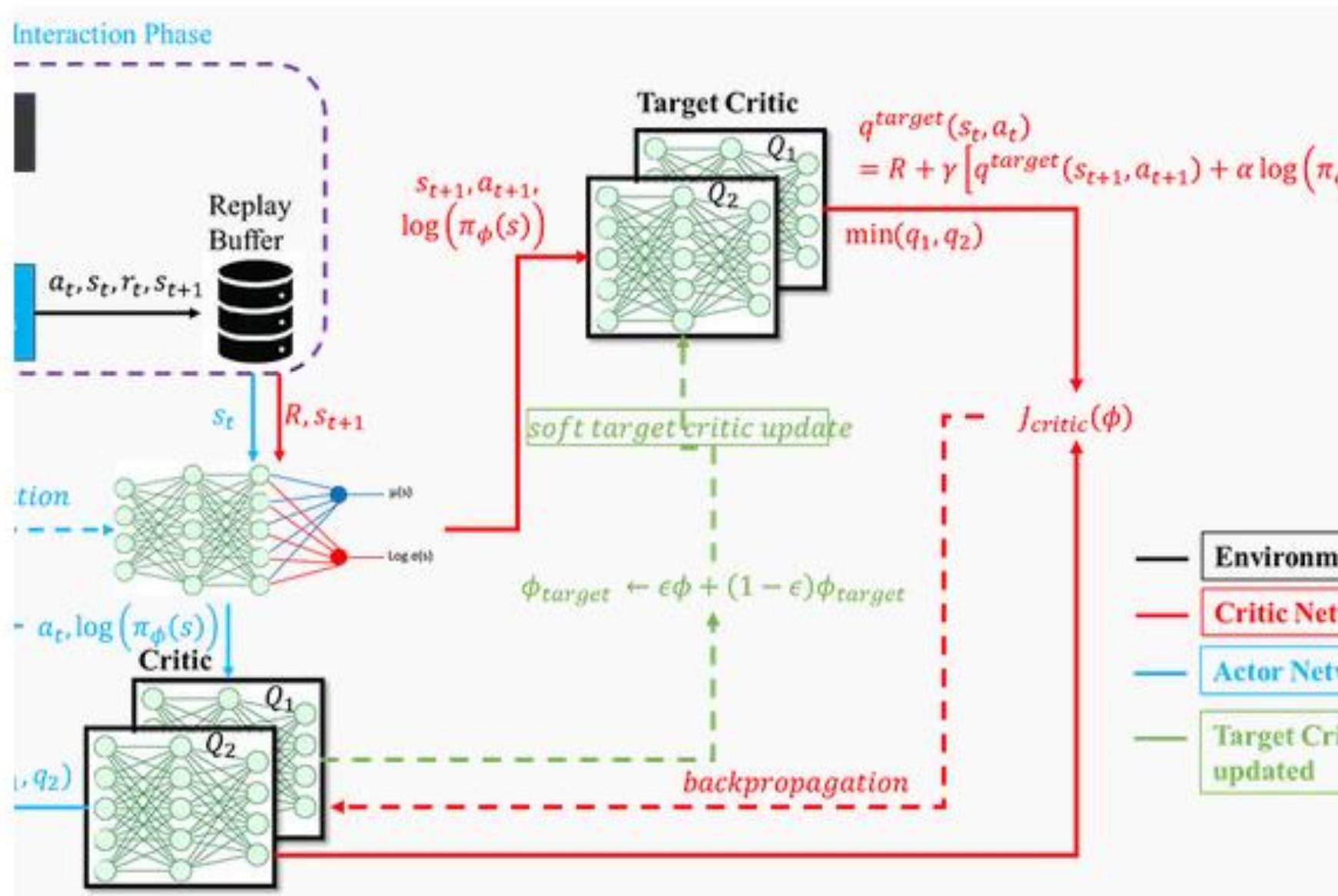
Re >

Machine Learning

Topic · 5.1M Followers · 361K Stories

Follow

Recommended stories



In AI Advances by Najib Sharifi, Ph.D.

Mathematical Foundation Underpinning



In The Generator by Jim the AI Whisperer

People considered “at risk” are more likely to get



Stay on top of the latest in AI

Get a 5-minute email summary of the latest breakthrough news, models, research, and repos.

Your best email...

GET MY SUMMARY

[Read the latest email](#)



Join 202.021 world-class developers from:



- Home
- Beliebt
- Erkunden
- Alle

INDIVIDUELLE FEEDS

- + Individuellen Feed erstellen

Zuletzt besucht

- r/MachineLearning

Communities

- + Erstelle eine Community
- r/MachineLearning ★
- r/OpenAI ★
- r/nlpclass ★

MEHR WISSEN

- Über Reddit
- Werben
- Hilfe
- Blog
- Karriere
- Presse
- Communities

r/MachineLearning

Heiß ▾ ▾

Community-Highlights

[D] Simple Questions Thread
0 Votes • 4 Kommentare

[D] Monthly Who's Hiring and Who wants to be Hired?
31 Votes • 2 Kommentare

u/alexsh1 vor 2 Std.

[R] Numerical features with factorization machines

Happy to share our recent [TMLR paper](#), "Function Basis Encoding of Numerical Features in Factorization Machines", by Alex Shtoff, Elie Abboud, Rotem Stram, and Oren Somekh. This paper proposes an interesting insight into the interplay between Factorization Machines (FMs), and feature encoding using basis functions, in the context of recommender systems. The same interplay with linear models is an old classic, and most of us have learned in our ML 101 courses. Polynomial regression is one of them - we encode a feature x using the standard polynomial basis $\{1, x, x^2, \dots\}$. FMs are family of models that model a quadratic polynomial...

Research

↑ 9 ↓ 0 Teilen

u/stonedoubt vor 12 Std.

[R]AST+Shorthand+HybridRag

Research

TEELINEAST: EFFICIENT CODE-SPECIFIC RAG THROUGH STRUCTURAL COMPRESSION AND MULTI-MODAL RETRIEVAL

Shawn McAllister
Engineered Automated Systems for Artificial Intelligence
Kannapolis, NC 28081
(Shawn McAllister) essaifounder@icloud.com

December 29, 2024

ABSTRACT

Current Retrieval-Augmented Generation (RAG) systems treat source code as natural language text, leading to inefficient storage, imprecise retrieval, and loss of structural information. We present TeelineAST, a novel approach that leverages Abstract Syntax Tree compression with multi-modal storage and retrieval mechanisms. Our theoretical framework demonstrates potential improvements of up to 90% in storage efficiency, 5x in retrieval speed, and 98% in accuracy compared to traditional approaches. TeelineAST introduces a paradigm shift in code-specific RAG systems by preserving

Machine Learning
Beginners -> /r/mlquestions , AGI -> /r/singularity, career advices -> /r/cscareerquestions, datasets -> r/datasets
Erstellt am 29. Juli 2009
Öffentlich

2,9 Mio. 128 Oberste 1 %
Mitglieder • Online Nach Größe sortieren

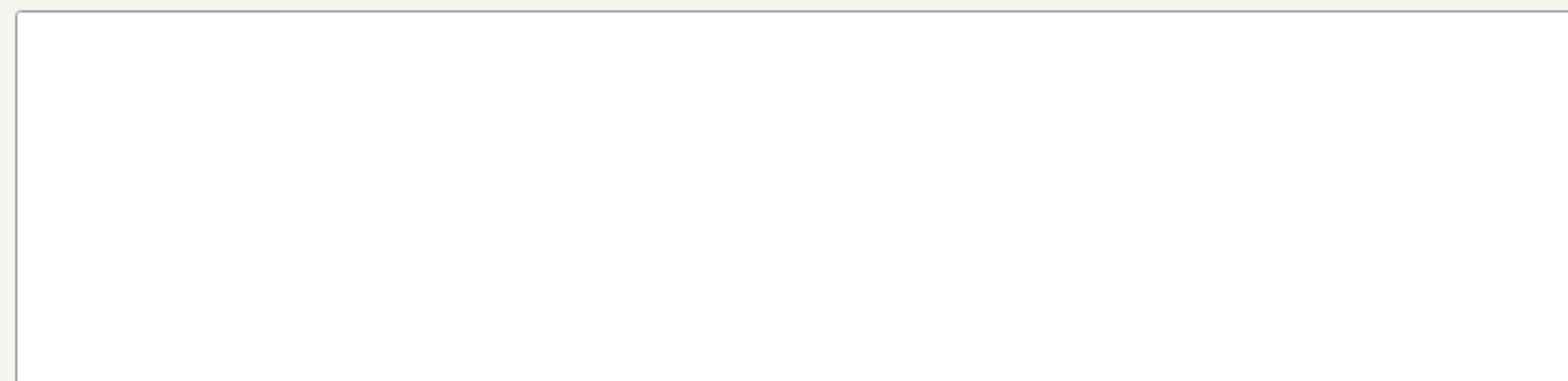
COMMUNITY-ERFOLGE

Langzeitmitglied
1 freigeschaltet Alle anzeigen

- REGELN
- 1 Be nice: no offensive behavior, insults or attacks
 - 2 Make your post clear and comprehensive
 - 3 Posts without appropriate tag in title will be removed
 - 4 Beginner or career related questions go elsewhere
 - 5 Non-arxiv link posts only allowed on weekends (must be demos)*
 - 6 Beginner's tutorials and projects go elsewhere
 - 7 Quality Contribution
 - 8 Limit self-promotion

▲ Safe Superintelligence Inc. (ssi.inc)

1023 points by nick_pou 16 hours ago | hide | past | favorite | 860 comments



[add comment](#)

▲ insane_dreamer 15 hours ago | next [-]

I understand the concern that a "superintelligence" will emerge that will escape its bounds and threaten humanity. That is a risk.

My bigger, and more pressing worry, is that a "superintelligence" will emerge that does not escape its bounds, and the question will be which humans control it. Look no further than history to see what happens when humans acquire great power. The "cold war" nuclear arms race, which brought the world to the brink of (at least partial) annihilation, is a good recent example.

Quis custodiet ipsos custodes? -- That is my biggest concern.

Update: I'm not as worried about Ilya et al as commercial companies (including formerly "open" OpenAI) discovering AGI.

[reply](#)

▲ benreesman 5 hours ago | parent | next [-]

It's just clearly military R&D at this point.

And it's not even a little bit controversial that cutting edge military R&D is classified in general and to an extreme in wartime.

The new thing is the lie that it's a consumer offering. What's new is giving the helm to shady failed social network founders with no accountability.

These people aren't retired generals with combat experience. They aren't tenured professors at Princeton IAS on a Nobel shortlist and encumbered by TS clearance.

They're godawful almost ran psychos who never built anything that wasn't extractive and owe their position in the world to pg's partisanship 15 fucking years ago.

[reply](#)

▲ yayr 0 minutes ago | root | parent | next [-]

most technology is dual or multiple use, starting with a rock or knife...

so it is up to the fabric of our society and everyone involved in dealing with the technology, how the rules and boundaries are set.

that there will be military use is obvious. However, it is naive to think one can avoid military use by others by not enabling oneself for it.

[reply](#)

▲ sgregnt 3 hours ago | root | parent | prev | next [-]

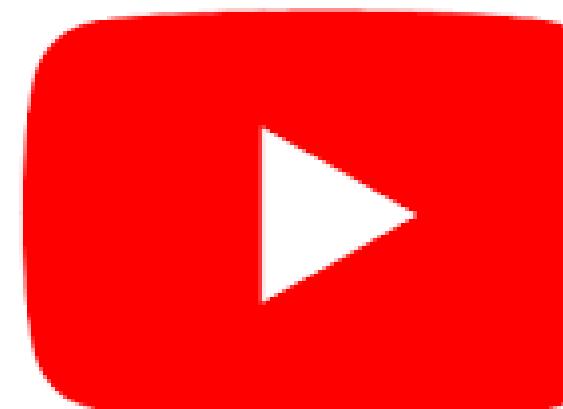
To me it is not clear at all, can you please elaborate why you make such a strong claim?

LINKEDIN

- **Philipp Schmid**
- **Lior Sinclair**
- **Hugging Face**
- **NVIDIA**
- ...



Udemy



YouTube

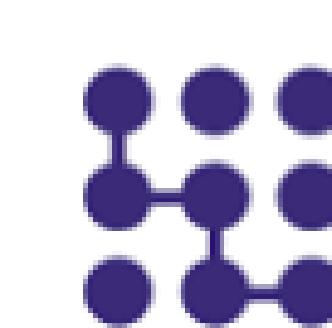
coursera

UDACITY

OPEN



Hasso
Plattner
Institut



KI-Campus

Die Lernplattform
für Künstliche Intelligenz



Machine Learning with TensorFlow

DONNERSTAG 18:15 - 20:00

Get hands-on experience in applying machine learning techniques with TensorFlow.

Die Bewerbungsfrist ist leider abgelaufen. Bleibe immer auf dem Laufenden und verpasste nicht unser nächstes Kursprogramm, Events, Jobs und was es sonst noch Neues gibt: [ZUM NEWSLETTER](#)

Du wirst lernen

- ✓ Mastering best practices for TensorFlow, a popular open-source machine learning framework, to train neural networks
- ✓ Handling real-world image data and exploring strategies to avoid overfitting, including techniques like augmentation and dropout
- ✓ Building a system capable of processing natural language
- ✓ Applying Recurrent Neural Networks (RNNs), Gated Recurrent Units (GRUs), and Long Short-Term Memory (LSTMs) networks to train models using text and time-series data

DONNERSTAG

18:15 -
20:00

ECTS

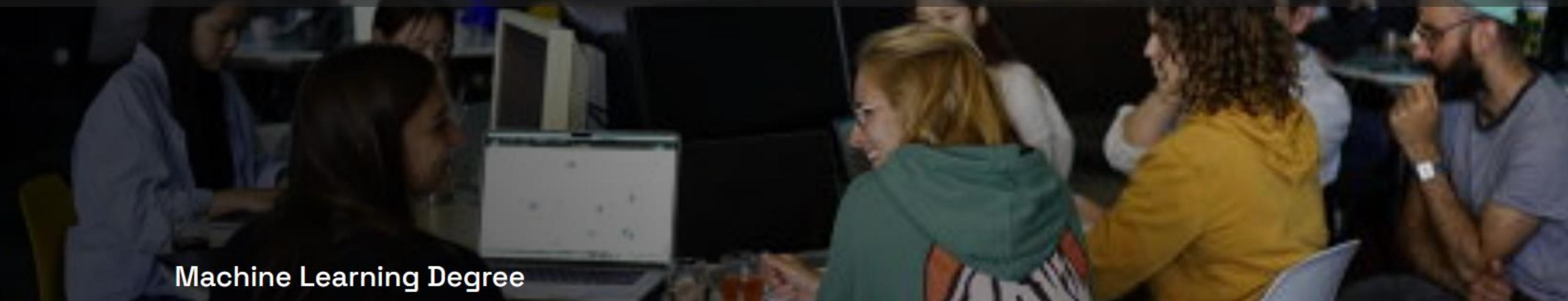
5



HEIDE +
KIEL +
ONLINE



ENGLISCH



Machine Learning Degree

After completing this program you will have a solid understanding of machine learning and will be able to implement your own state of the art machine learning projects.

Die Bewerbungsfrist ist leider abgelaufen. Bleibe immer auf dem Laufenden und verpasste nicht unser nächstes Kursprogramm, Events, Jobs und was es sonst noch Neues gibt: [ZUM NEWSLETTER](#)

Aktuelle Angebote im Rahmen des Degrees

- [Advanced Time Series Prediction](#)
- [Intermediate Machine Learning](#)
- [AI Builder's Arena](#)
- [Machine Learning with TensorFlow](#)
- [Einführung in Data Science & maschinelles Lernen](#)
- [TinyML: Embedded Machine Learning](#)

ECTS

12,5

KIEL +
ONLINE

ENGLISCH



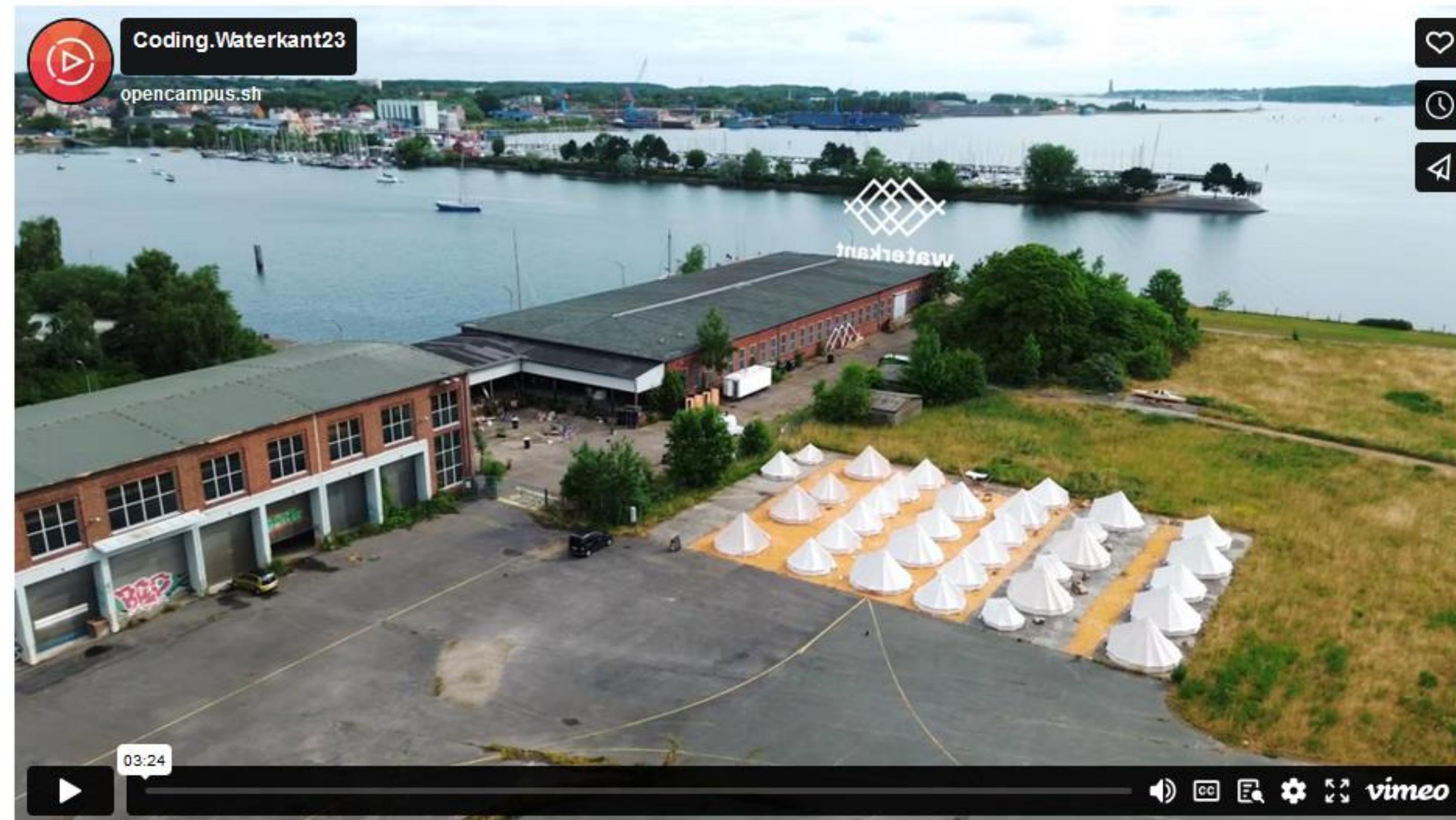
Mehr

CODING.WATERKANT 2025

Save the Date

07.07.-11.07.2025

Um vom Registrierungsstart zu erfahren und von den Early-Bird-Preisen zu profitieren,
[melde Dich hier an.](#)





Coding.Waterkant

★★★★★ (128) 2

📍 Kiel, Deutschland

👤 1.003 Mitglieder · Öffentliche Gruppe i

👤 Organisiert von [opencampus.sh](#) and **6 others**

✉️ [Mitglieder kontaktieren](#)

Teilen: [n](#) [f](#) [t](#) [in](#) [✉](#)

Info

Events

Mitglieder

Fotos

Diskussionen

Mehr

Event erstellen ▼

Gruppe verwalten ▼

Über uns

Our meetup is organized with support of [opencampus.sh](#) and the Digitale Wirtschaft Schleswig-Holstein (DiWiSH).

[Mehr lesen](#)



Organizers



[opencampus.sh and 6 others](#)

[✉ Nachricht](#)

Members (1.003)

[Alles ansehen](#)

