

20.06.24

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

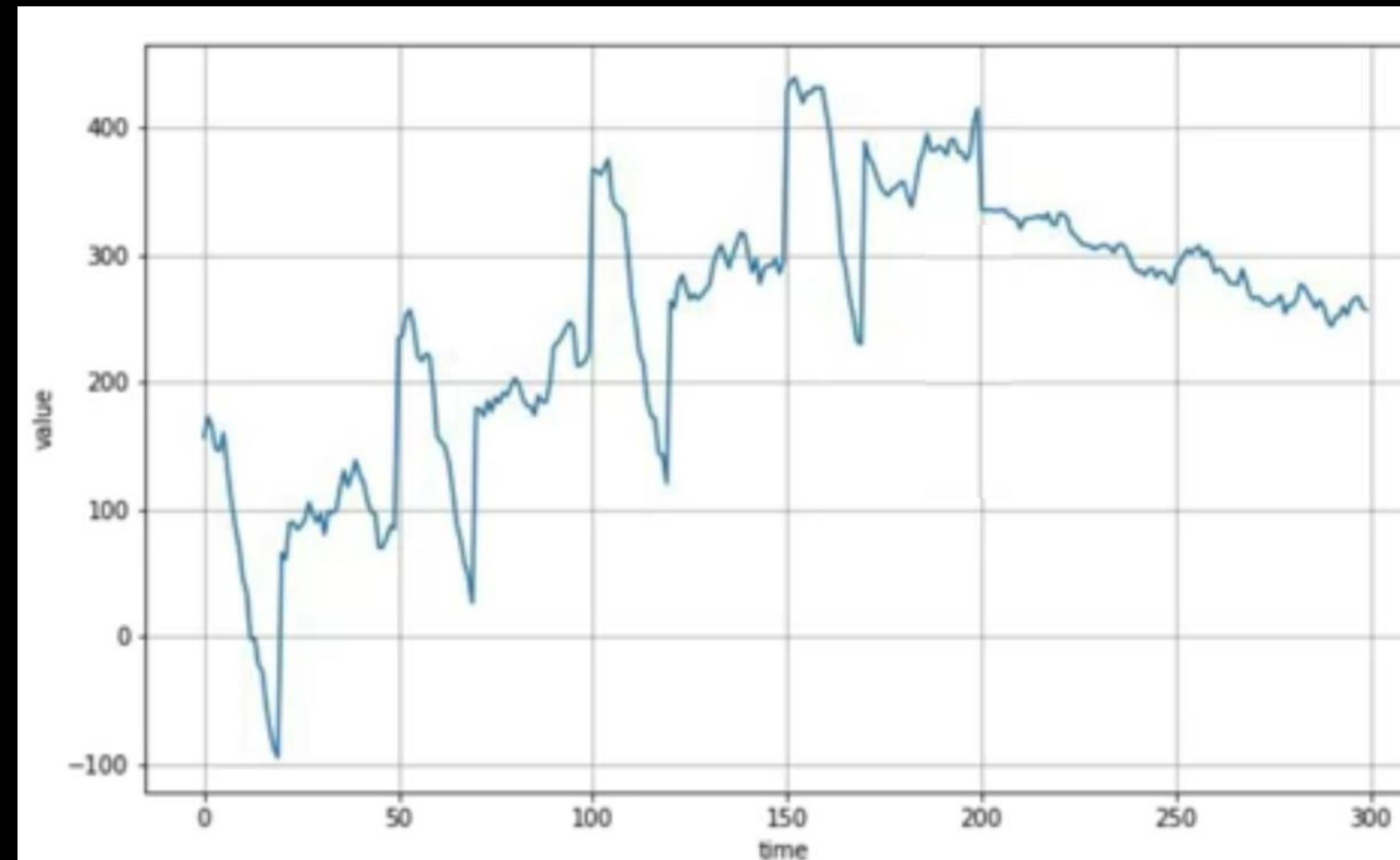
- **Betrachtung der Auswirkung auf die Vorhersage:
Berechnung der Vorhersagen für den Validierungsdatensatz
mit und ohne Imputationen**
- **Bewerten der Genauigkeit von imputierten Werten im
Vergleich zu bekannten Originalwerten:
Erstellen künstlicher Fehlwerte und Anwenden der
Imputationsmethode**

IMPUTATION BEI VIELEN FEHLENDEN WERTEN

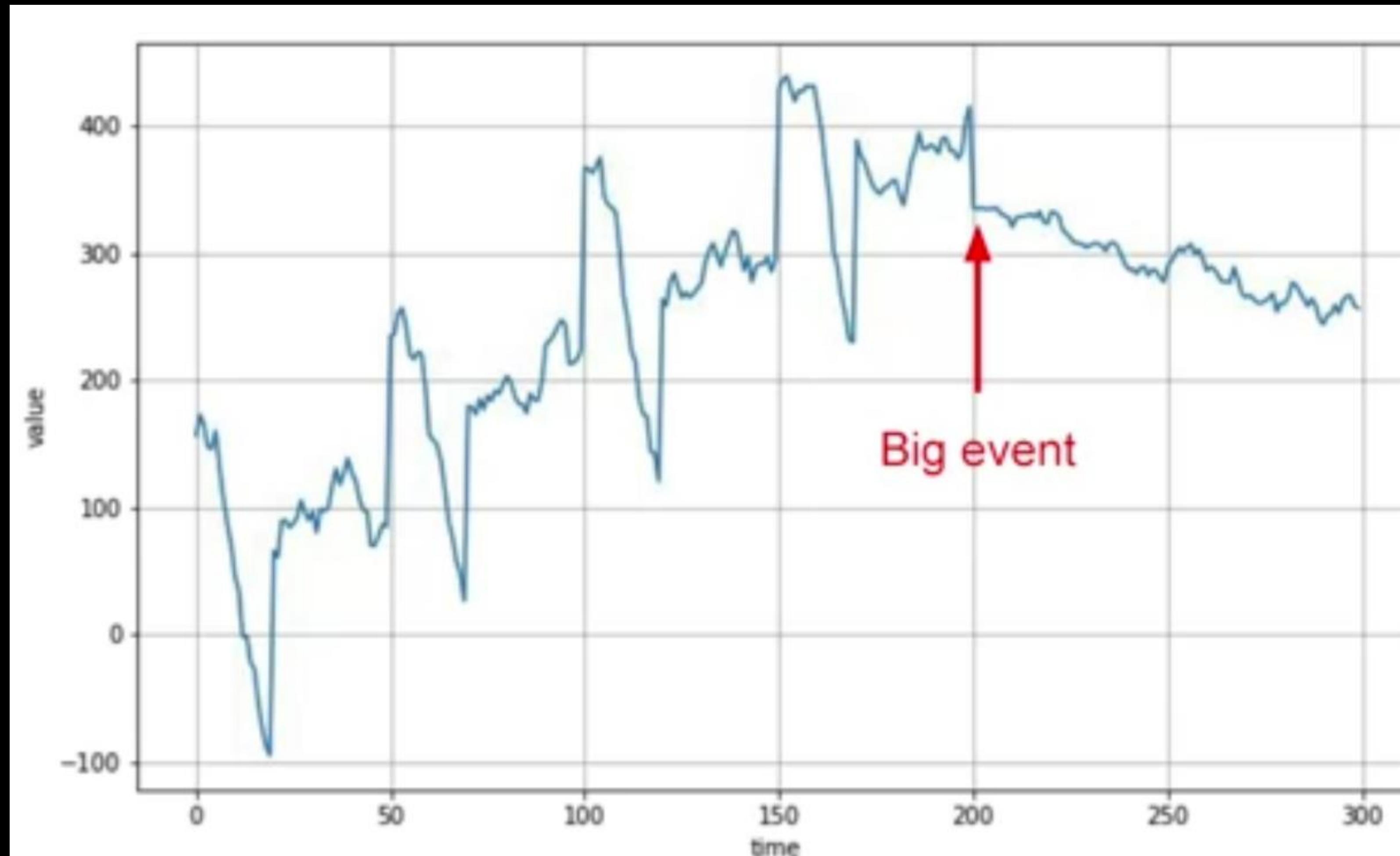
- Berücksichtigung der Information, welche Werte imputiert wurden, als Feature.

DISKUSSION

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



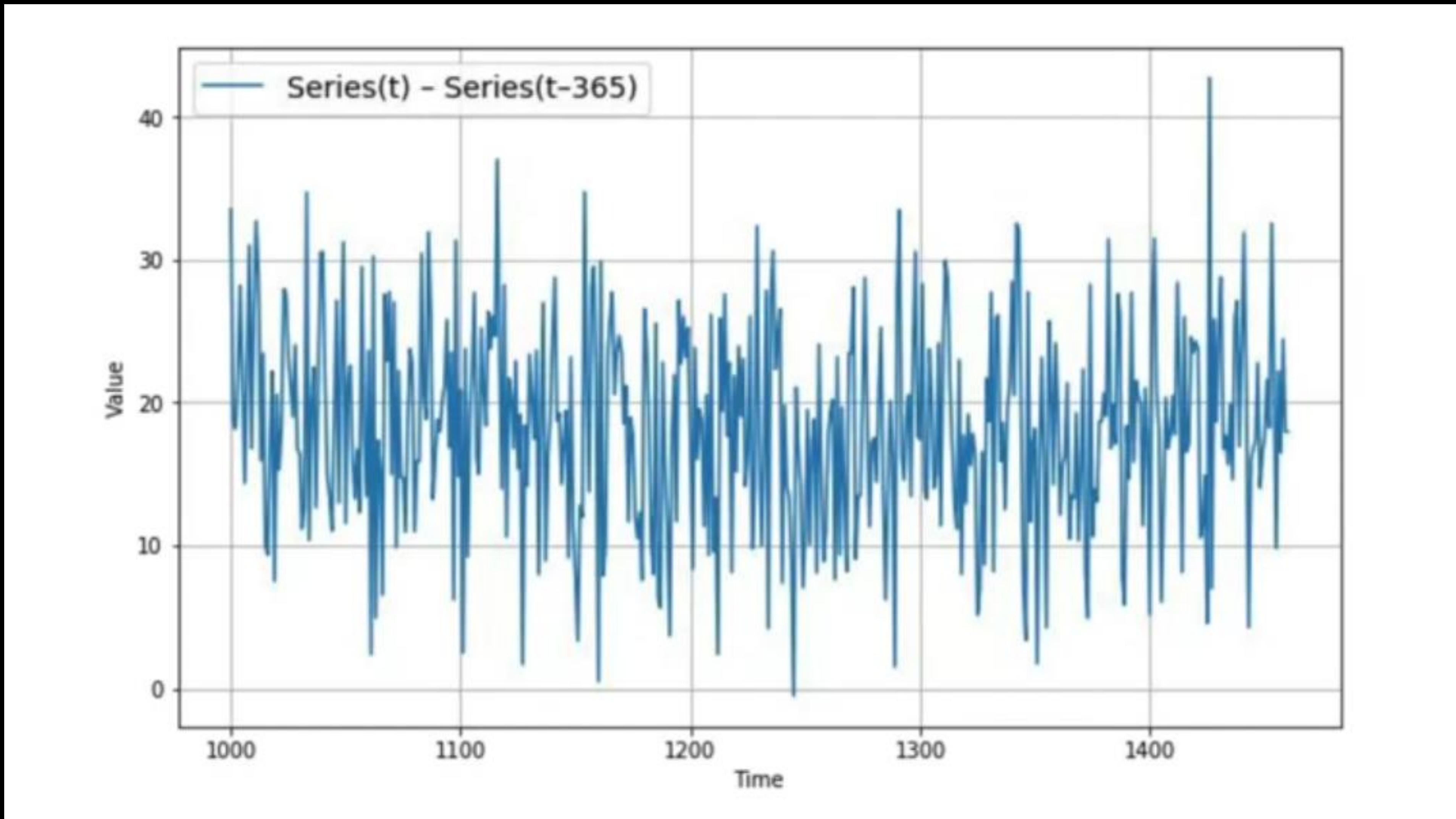
NON-STATIONARITY



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 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

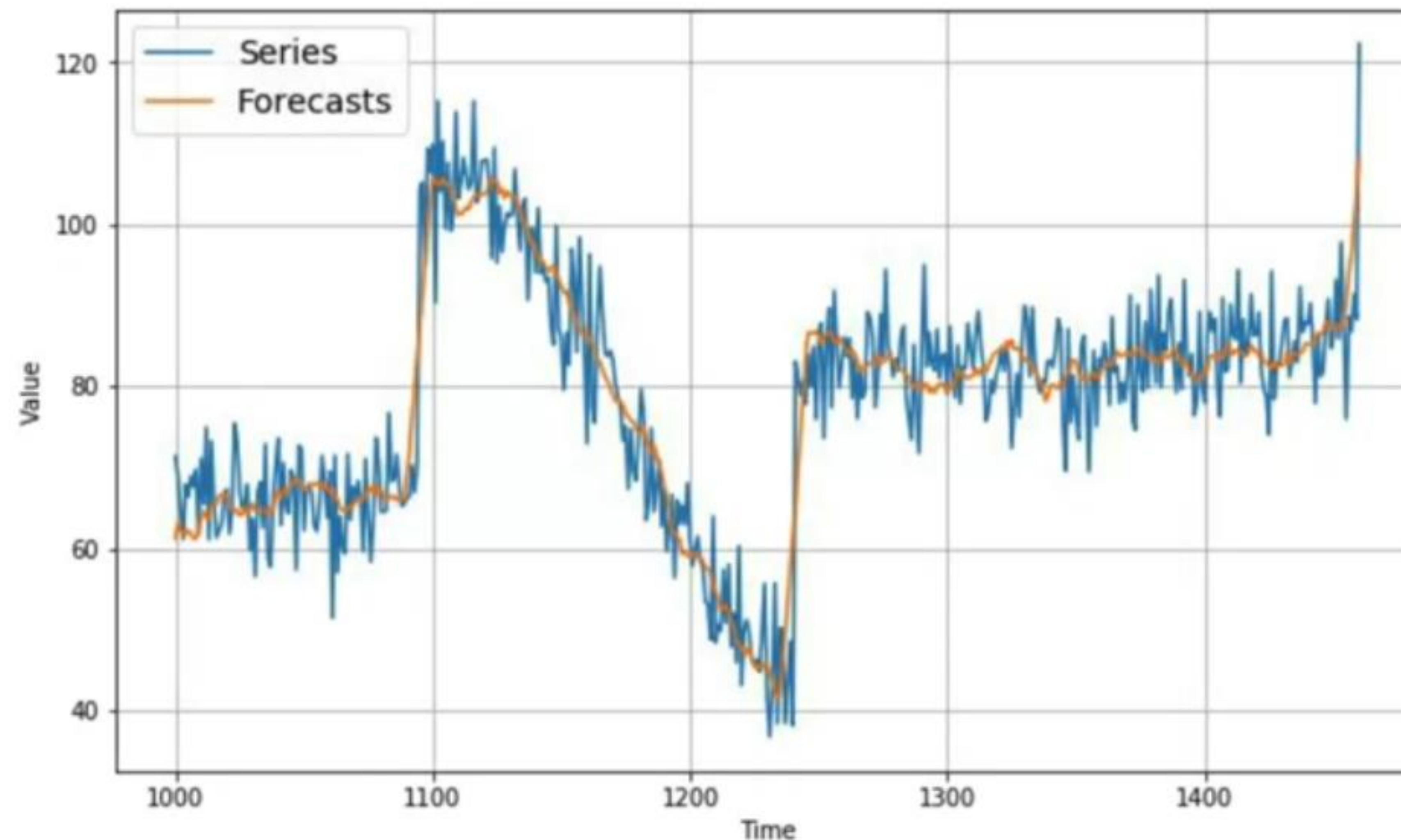
- Allein an Metriken wie MAPE oder RMSE kann man häufig schlecht abschätzen, wie viel das eigene Modell gelernt hat.
- Es ist daher wichtig, die Ergebnisse anderer Modelle als „Baseline“ bzw. Referenz zu nutzen.

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 ($t - 365$)

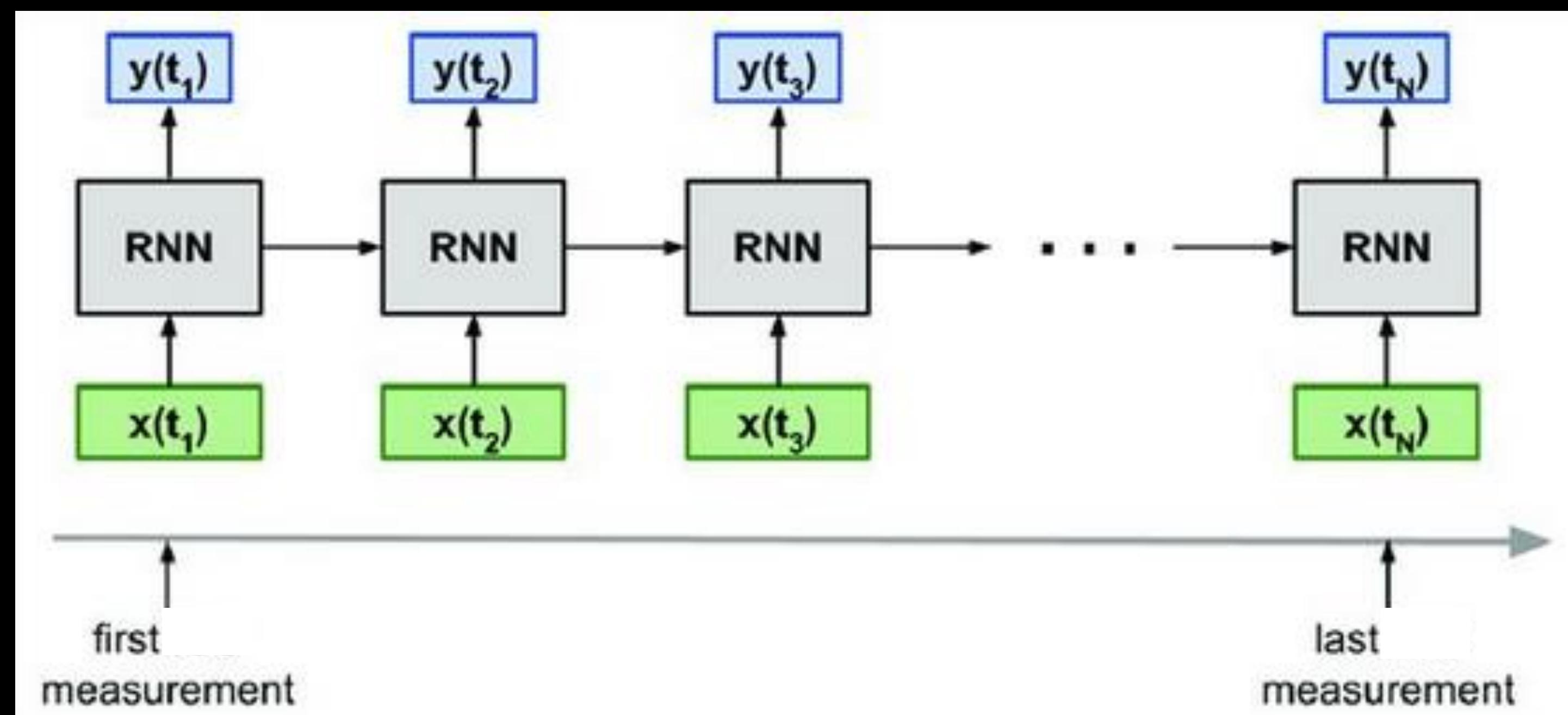
DISKUSSION

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

- Welche Informationen aus der Zeitreihe wären zusätzlich 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. August:

- **Die Präsentation zu Eurem Repo hinzufügen.**
- **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 in der Praxis

6. Juni 2023

Apple spricht nicht über KI, packt sie aber trotzdem überall rein



DEEP MINDS Podcast

Podcast über Künstliche Intelligenz und Wissenschaft



Künstliche Intelligenz und Robotik | DEEP MINDS #15

Verfügbar bei [Youtube](#), [Soundcloud](#), [Spotify](#), [Apple](#), [Google](#) und [Amazon](#)





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

Replies to @b0rk
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





Search Medium

Write



Get unlimited access to all of Medium. [Become a member](#)



Machine Learning

Follow

Start writing

Trending Latest Best

 Merve Noyan · 17 hours ago

Complete Guide on Deep Learning Architectures Part 2: Autoencoders

Autoencoder: Basic Ideas Autoencoder is the type of a neural network that reconstructs an input from the output. The basic idea here is tha...



Machine Learning 5 min read



...

 Bex T. in Towards Data Science · 18 hours ago ✨ Member-only

10 Confusing XGBoost Hyperparameters and How to Tune Them Like a Pro in 2023



251K

Stories

92K

Writers



Related Topics

Data Science

Artificial Intelligence

Deep Learning

Python

AI

Technology

Programming

NLP

Neural Networks

[See more topics](#)

Top Writers



The PyCoach

10M+ Views on Medium II

Follow

AlphaSignal

Hey Steffen,

Welcome to today's edition of AlphaSignal.

IN TODAY'S SIGNAL

💡 Top News

- [DeepSeek-Coder-V2](#) outperforms GPT4-Turbo in coding, math benchmarks

📌 Latitude

- [Deploy AI applications](#) instantly with Launchpad – Purpose-built for speed and scalability on dedicated GPUs.

⚡ Top 5 Signals

- DeepMind releases V2A model: generates [accurate video soundtracks](#)
- [Open-Sora v1.2](#): produces high-quality video, open-source way
- Gemini API introduces [context caching](#), reduces costs by 50%
- Retool releases AI report: insights from [700+ tech leaders](#)
- ElevenLabs unveils [text-to-sound effects API](#)

📄 Top Papers

- [Monte Carlo Trees](#) + Llama-3 8B solves mathematics limitations of LLMs
- Samba architecture achieves [3.73x faster throughput](#) with enhanced context
- Depth Anything V2 model: delivers 10x faster, more [accurate depth predictions](#)

🧠 Recommended Lecture

- [Step-by-Step Diffusion](#): Comprehensive Tutorial on Diffusion Models



reddit



r/MachineLearning

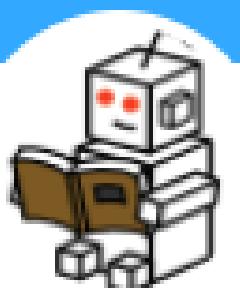


r/MachineLearning

Reddit durchsuchen



Gratis

matanzino
1 Karma

Machine Learning

r/MachineLearning

[Beigetreten](#)[Beiträge](#)

Beitrag erstellen



Heiß



Neu



Top



...



VON EINEM MOD ANGEHEFTET

6

Gepostet von u/AutoModerator vor 1 Tag



Discussion [D] Simple Questions Thread



20 Kommentare



Auszeichnen



Teilen



Merken



Gepostet von u/ML_WAYR_bot vor 22 Stunden

4

Discussion [D] Machine Learning - WAYR (What Are You Reading) - Week 128



1 Kommentar



Auszeichnen



Teilen



Merken



Gepostet von u/jayalammar vor 5 Stunden



120

Research [R] The Illustrated Retrieval Transformer (GPT3 performance at 4% the size)

Hi r/MachineLearning,

Über diese Community

...

Willkommen in MachineLearning

2.2m

Mitglieder

488

Online

Am 29. Juli 2009 erstellt

[Beitrag erstellen](#)

COMMUNITY-EINSTELLUNGEN

Nach Flair filtern

Discussion

Research

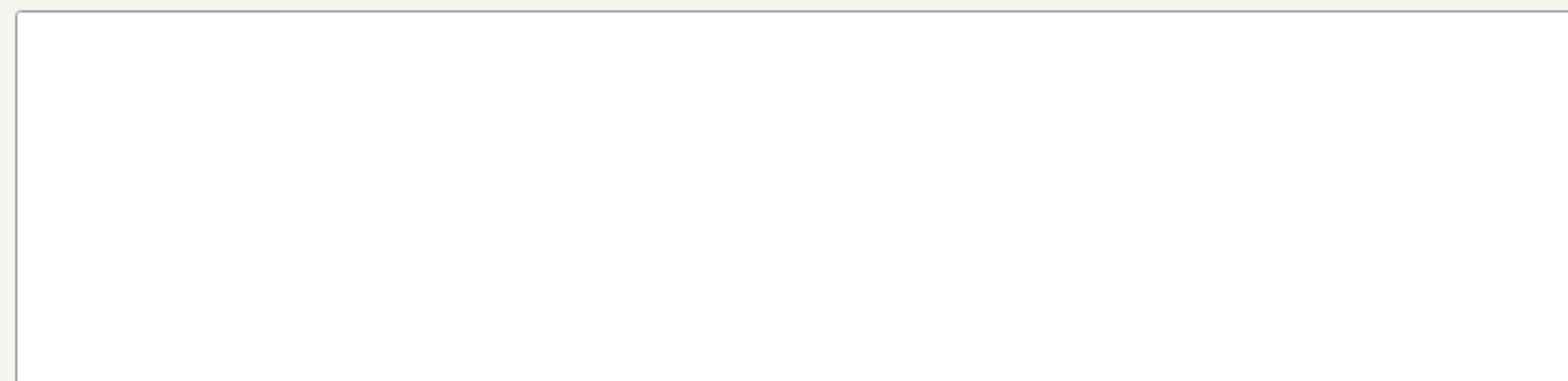
Project

Twitter

@clashML

▲ 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?

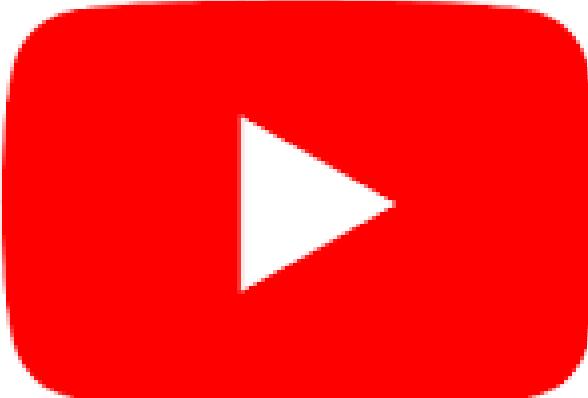
LINKEDIN

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



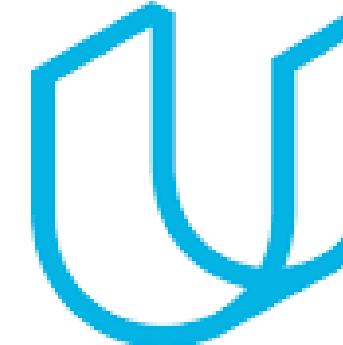
 Udemy

The Udemy logo features a stylized red "u" character followed by the word "Udemy" in a dark grey, lowercase, sans-serif font.

 YouTube

The YouTube logo consists of a red rounded rectangle containing a white play button icon, followed by the word "YouTube" in a large, black, sans-serif font.

coursera

 UDACITY

The Udacity logo features a blue stylized "U" character followed by the word "UDACITY" in a dark grey, uppercase, sans-serif font.

 Hasso
Plattner
Institut

The OPEN HPI logo features the word "OPEN" in large orange letters, followed by "HPI" in white inside a yellow square, and "Hasso Plattner Institut" in a smaller dark grey font.

 | KI-Campus
Die Lernplattform
für Künstliche Intelligenz

The KI-Campus logo features a purple graphic of connected dots on the left, a vertical line, and the word "KI-Campus" in a large purple font. Below it, the text "Die Lernplattform für Künstliche Intelligenz" is written in a smaller purple font.

{ MACHINE # LEARNING > DEGREE }

BY OPENCAMPUS.SH

GET A SOLID UNDERSTANDING OF
MACHINE LEARNING AND LEARN
HOW TO IMPLEMENT YOUR OWN
STATE OF THE ART MACHINE
LEARNING PROJECTS.

PARTICIPATE AS INDIVIDUAL:

→ REGISTER FOR OUR NEXT INFO
EVENT

BENEFIT AS AN ORGANIZATION:

→ SUBMIT A COURSE PROJECT
PROPOSAL

{ }

#

>

Machine Learning with TensorFlow

DIENSTAG 16:00 - 17:45

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

Die Bewerbungsfrist ist leider abgelaufen.

Du wirst lernen

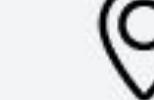
- ✓ Best Practices für TensorFlow, ein populäres Open-Source-Framework für maschinelles Lernen, um neuronale Netzwerke zu trainieren
- ✓ Umgang mit Bilddaten aus der realen Welt und Erkundung von Strategien zur Vermeidung von Overfit, einschließlich Augmentation und Drop-Out
- ✓ Erstellung eines Systems zur Verarbeitung natürlicher Sprache
- ✓ Anwendung von RNNs, GRUs und LSTMs zum Training dieser Lernmodelle unter Verwendung von Text- und Zeitreihendaten

DIENSTAG

16:00 -
17:45

ECTS

5



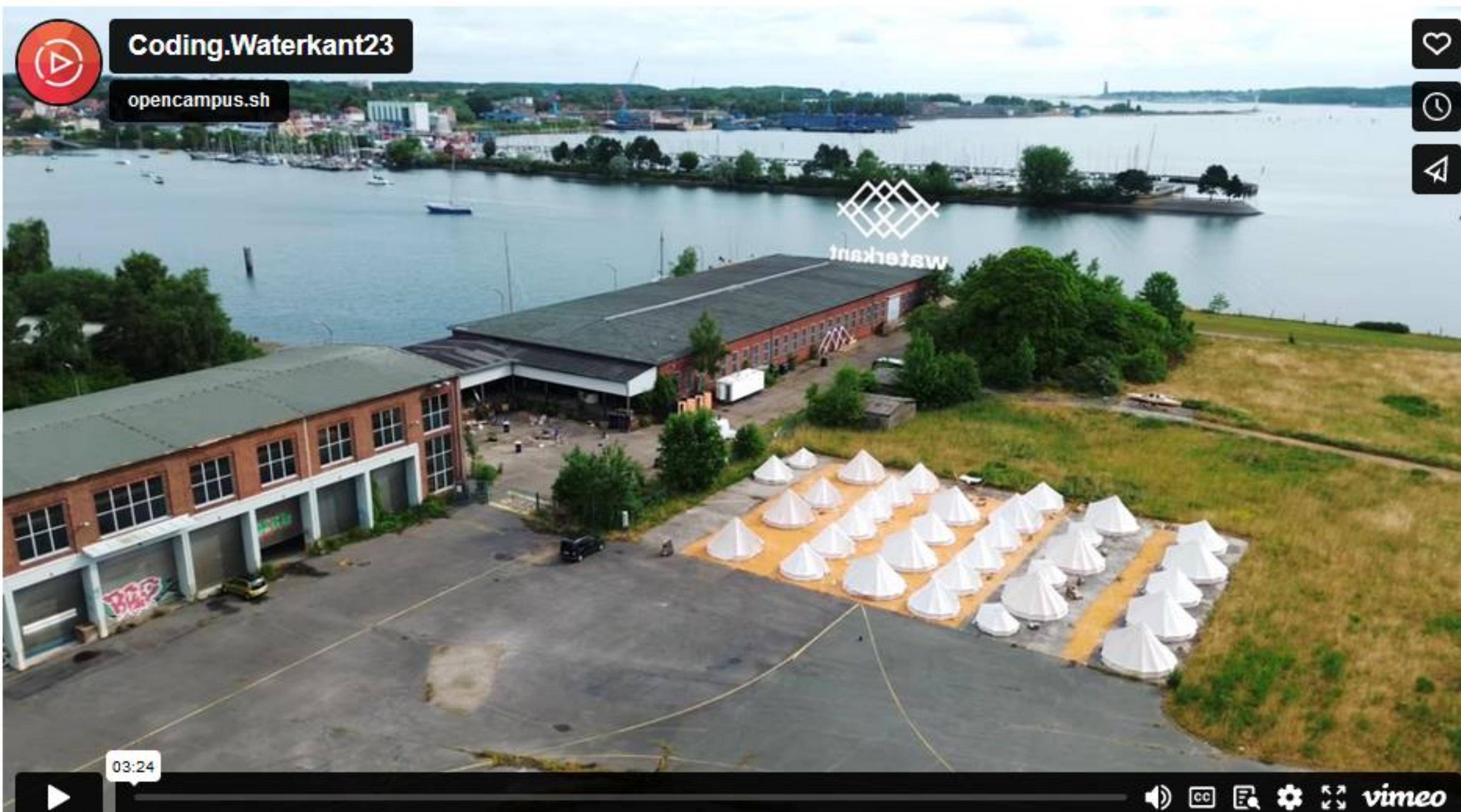
ONLINE +
KIEL



ENGLISCH

CODING.WATERKANT 2024

01.07.- 05.07.2024



[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](#)

Coding.Waterkant

★★★★★ (117)

Neu: Event-Feedbackübersicht

Tippe auf die Sterne, um alle deine Eventbewertungen in der Übersicht anzuzeigen.

[Weitere Informationen](#)

Kiel, Deutschland

948 Mitglieder · Öffentliche Gruppe

Organisiert von **opencampus.sh** and **6 others**

Teilen:

Organizers



opencampus.sh and 6 others

[Nachricht](#)

Members (948)

[Alles ansehen](#)

Du nimmst an diesem Event teil!

Christoph Schuhmann: LAION and its Role in Open Source AI



Hosted By
Steffen und 2 weitere



Details

We have exciting news: Kiel.AI is now Coding.Waterkant!

Since 2020, we have been organizing Coding.Waterkant, an event dedicated to implementing machine learning projects. By renaming the Meetup, we aim to bring the

Organisatoren-Tools ▾



Coding.Waterkant
Public group

- Montag, 1. Juli 2024
18:00 bis 19:00 MESZ
[Zum Kalender hinzufügen](#)
- Waterkant Festival
Halle 51 · Kiel, SH
So findest du uns
Waterkant Festival Schusterkrug 25,
Halle 51 24159 Kiel

