

Einführung in Data Science und maschinelles Lernen

ZEITREIHENANALYSEN

- ChatGPT als Alternative zu Google
- Muster in Zeitreihenanalysen
- Non-Stationarity
- Baseline Modelle und Naïve Forecasting
- Projektpräsentation
- Undjetzt?

CHATGPT

- Sogenanntes Large Language Model (LLM)
- Modelle, die üblicherweise "unsupervised" mit Textdaten trainiert werden.

 Weiterentwicklung von GPT3, das 175 Milliarden Parameter hat.

 ChatGPT wurde erweitert um ein Reinforcement Training basierend auf von Menschen gegebenen Beurteilungen.

VERWENDUNGSMÖGLICKEITEN

- Generierung von Inhaltsangaben
- Generierung von Zusammenfassungen
- Erklären schwieriger Konzepte
- • •
- Unterstützung beim Programmieren
- • •

MÖGLICHKEITEN IM BEREICH DER PROGRAMMIERUNG

- 1. Explaining code
- 2. Improve existing code
- 3. Rewriting code using the correct style
- 4. Rewriting code using idiomatic constructs
- 5. Simplifying code

- 6. Writing test cases
- 7. Exploring alternatives
- 8. Translating code
- 9. Writing documentation
- 10. Tracking down bugs
- 11. Scaffolding

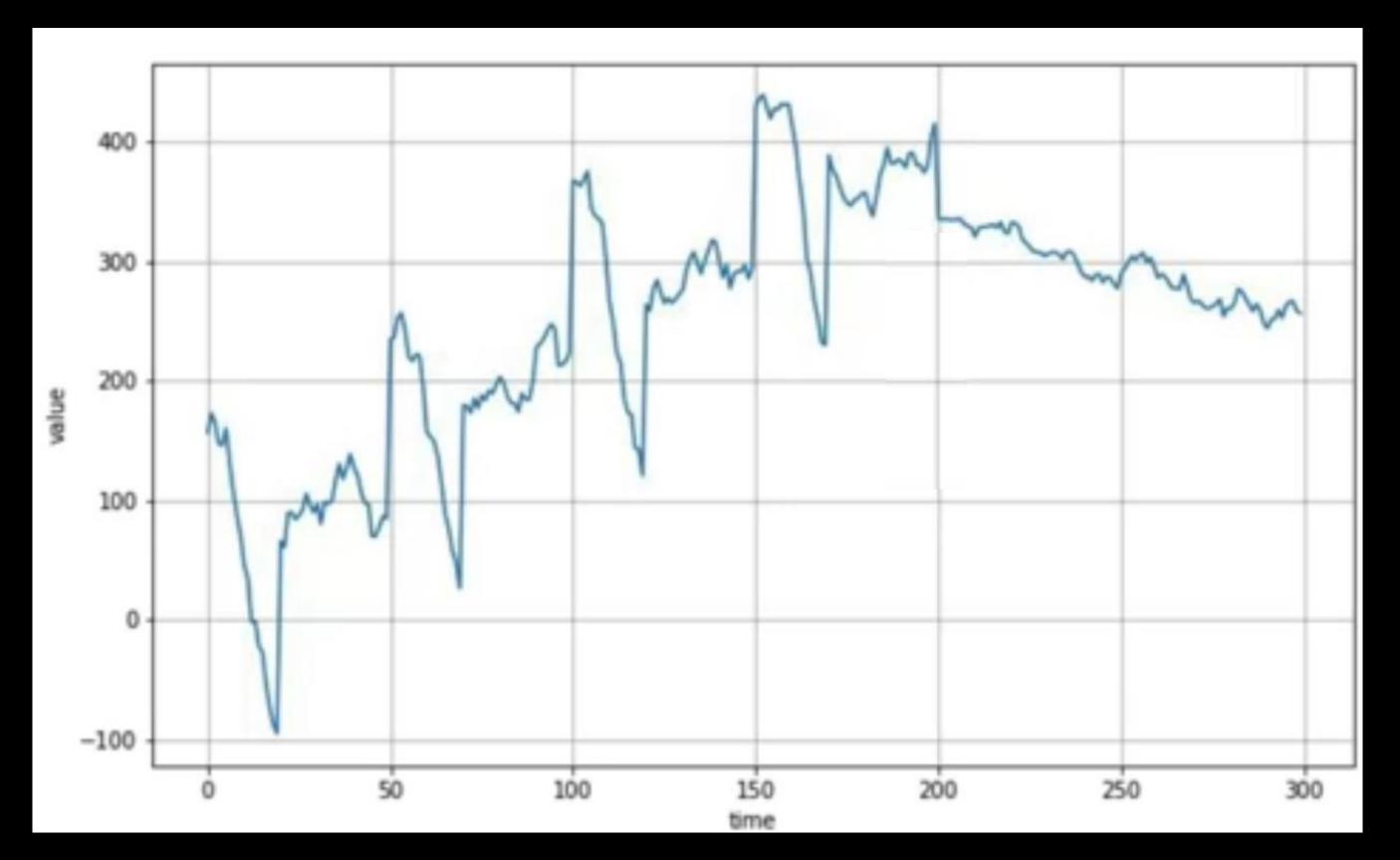
ZUR LERNRATE DES OPTIMIZERS

 Große Lernrate: Schnelle Optimierung, aber Gefahr, dass nicht das tatsächliche Minimum gefunden wird.

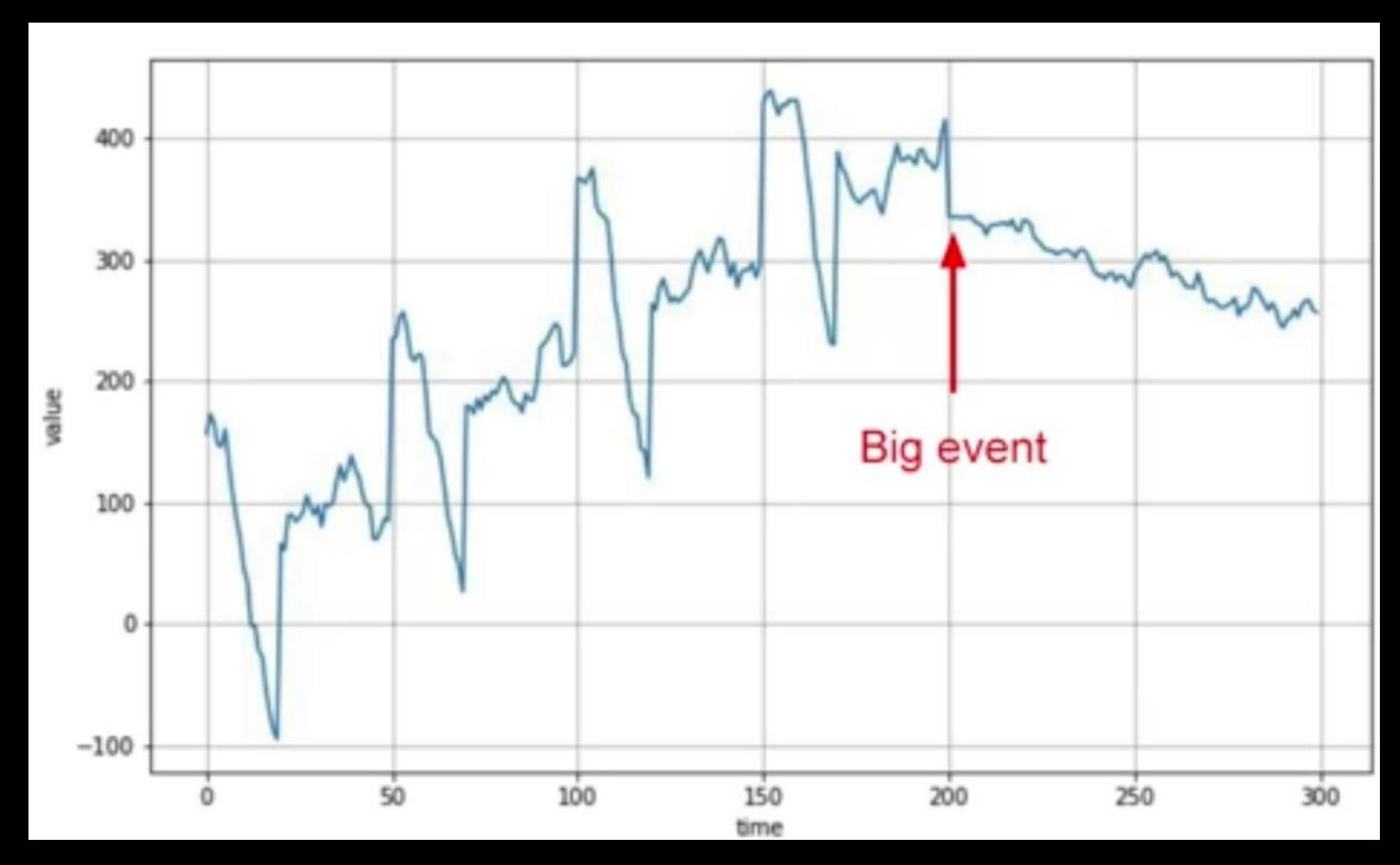
 Kleine Lernrate: Sehr langsame Optimierung, aber größere Sicherheit, dass man das tatsächliche Minimum findet.

DISKUSSION

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



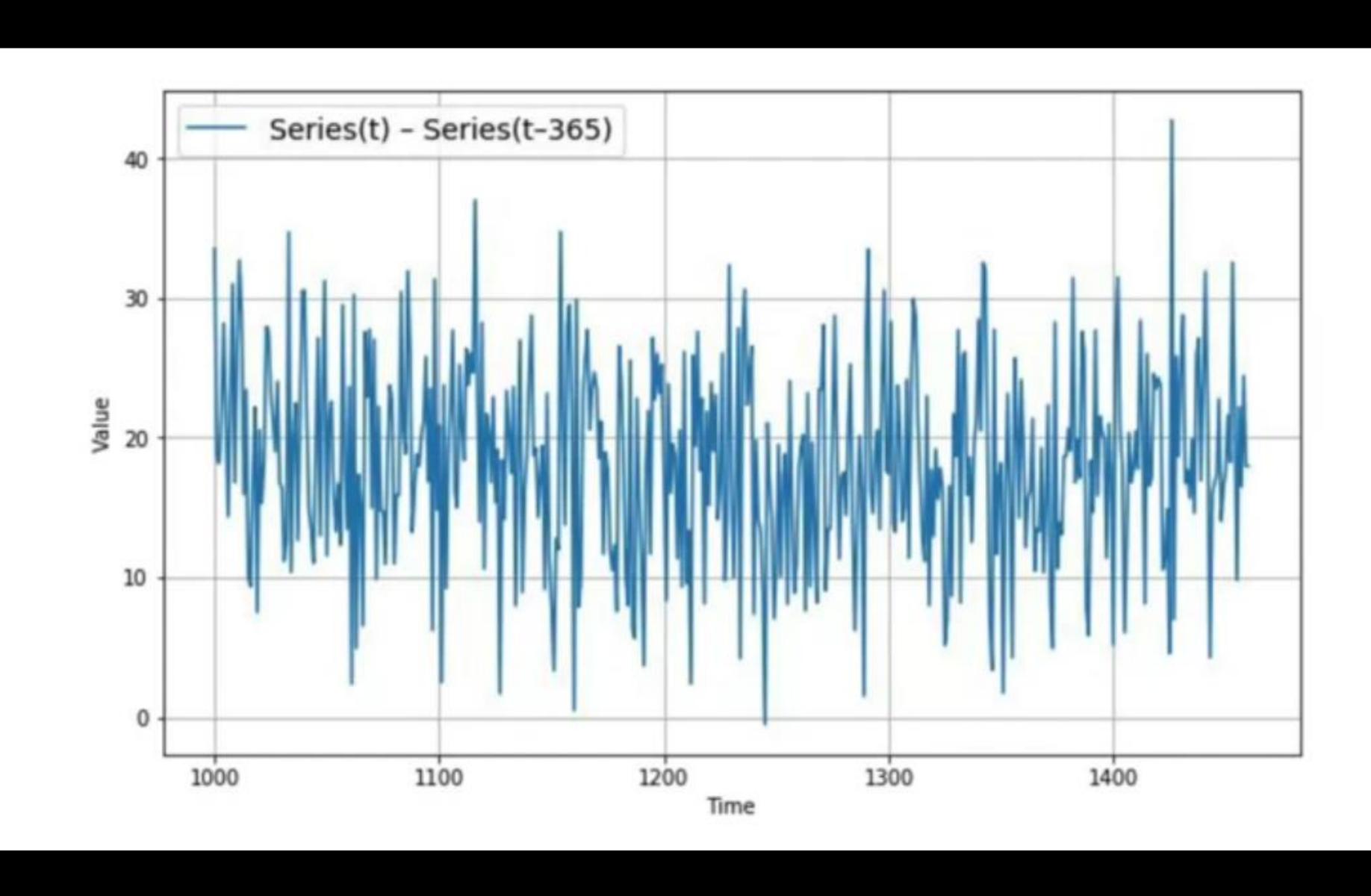
NON-STATIONARITY



MUSTER IN ZEITREIHENANALYSEN

- Trends
- "Jahresgang" (Saisonality)
- Rauschen (Noise)
- Autokorrelation

DIFFERENCING



DIFFERENCING

Subtraktion der vorherigen Beobachtung von der aktuellen Beobachtung

 Methode um die Zeitabhängigkeit von Zeitreihen zu entfernen bzw. zu mindern.

Versuch Zeitreihen "stationary" zu machen

LAG-DIFFERENZ

 Die Differenz zwischen aufeinanderfolgenden Beobachtungen wird als Lag-1-Differenz bezeichnet.

 Die Lag-Differenz kann an die spezifische zeitliche Struktur angepasst werden.

 Bei Zeitreihen mit einer saisonalen Komponente kann man davon ausgehen, dass die Verzögerung der Periode (Breite) der Saisonalität entspricht.

BASELINE MODELLE

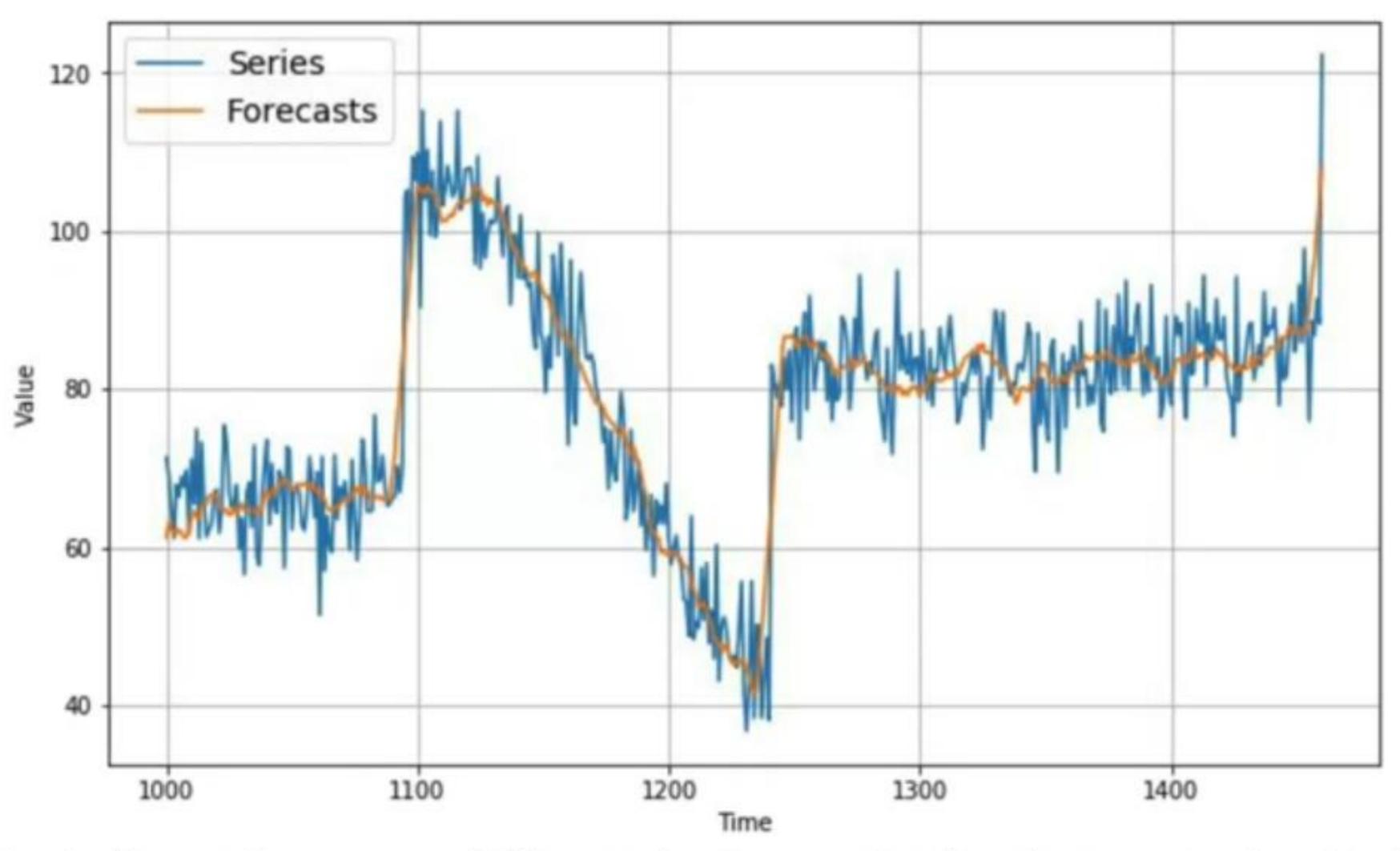
- Allein an Metriken wie MAPE oder RMSE kann man häufig schlecht abschätzen, wie viel das eigene Modell gelernt hat.
- Es ist daher wichtig, die Ergebnisse anderer Modelle als "Baseline" bzw. Referenz zu nutzen.

Mögliche Baselines:

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

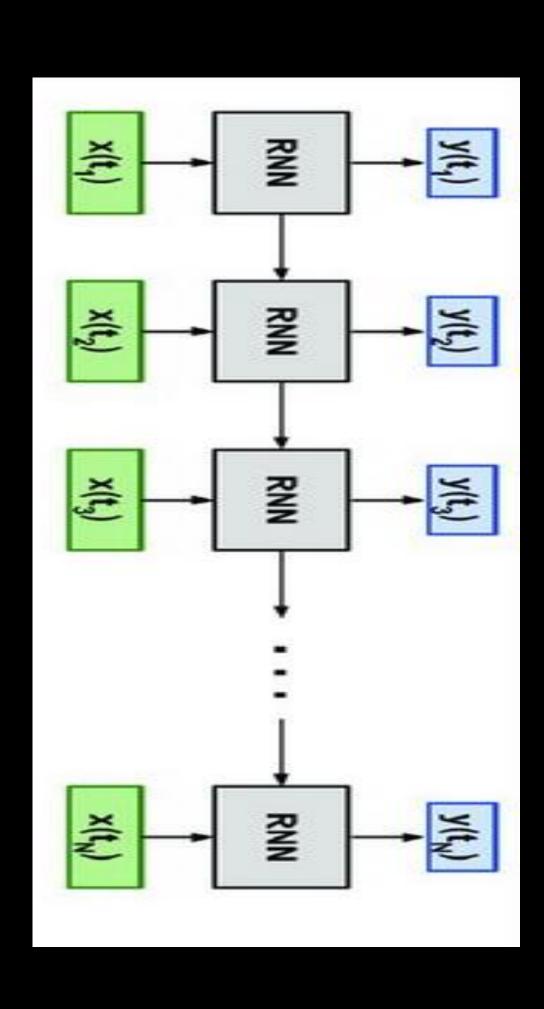
NAIVE FORECASTING

- Mögliches Baseline Modell für Zeitreihenanalysen
- Vorhersage entspricht dem jeweils letzten beobachteten Wert
- Saisonal 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)

RECURRENT NEURAL NETS



 Neurales Netz bei dem die Knoten einer Schicht untereinander verknüpft sind.

INHALT DER PROJEKTPRÄSENTATION

Dauer: ca. 8 Minuten pro Team

- Eure Namen
- Auflistung und kurze Beschreibung der selbst erstellten Variablen
- Balkendiagramme mit Konfidenzintervallen für zwei selbst erstellte Variablen
- Optimierung eines neuronalen Netzes:
 - Source Code
 - MAPE für einen Testdatensatz mit 10% der Ursprungsdaten
 - Warengruppenumsätze für den ersten nicht im Datensatz enthaltenen Tag

PRASENTATION

- Eure Namen auf der Titelseite
- Auflistung und kurze Beschreibung der selbst erstellten Variablen
- Balkendiagramme mit Konfidenzintervallen für zwei selbst erstellte Variablen
- Optimierung eines neuronalen Netzes
- Source code
- Das Evaluationsergebnis für den Zeitraum vom 9.5. bis 30.7.2019 (anhand der gegebenen Funktion \rightarrow Gruppennamen verwenden)

HINWEISE

- Nutzt ein formatiertes R-Markdown Notebook für die Präsentation
- Formatierung des Codes mithilfe von styler
- Erzählt gerne auch, was Ihr versucht habt, aber nicht geklappt hat.





R news and tutorials contributed by hundreds of R bloggers

HOME

ABOUT RSS ADD YOUR BLOG!

LEARN R

R JOBS

CONTACT US

How to perform Eta Squared in R

January 3, 2022 | finnstats

The post How to perform Eta Squared in R appeared first on finnstats. If you want to read the original article, click here How to perform Eta Squared in R. Eta Squared in R, Eta squared is a commonly-used effect size metric in ANOVA models. It is calculated as follows: ... [Read more...]

Uncovered Interest Rate Parity and F-test on Regression Parameters using R

January 2, 2022 | sang-heon lee

This post explains how to perform the F-test of joint parameter restrictions on a linear regression model. As an example, we use the data in Chen and Tsang (2013), who introduce so called relative Nelson-Siegel factor model to predict exchange r...

[Read more...]



Using databases with Shiny Examining College Football Conference

Most viewed posts (weekly)

How to perform the Sobel test in R

January 1, 2022 | finnstats

The post How to perform the Sobel test in R appeared first on finnstats. If you want to

Search R-bloggers.. Go

Your e-mail here

Subscribe

52793 readers

DuckDuckGo

DuckDuckGo blocked this Facebook Page

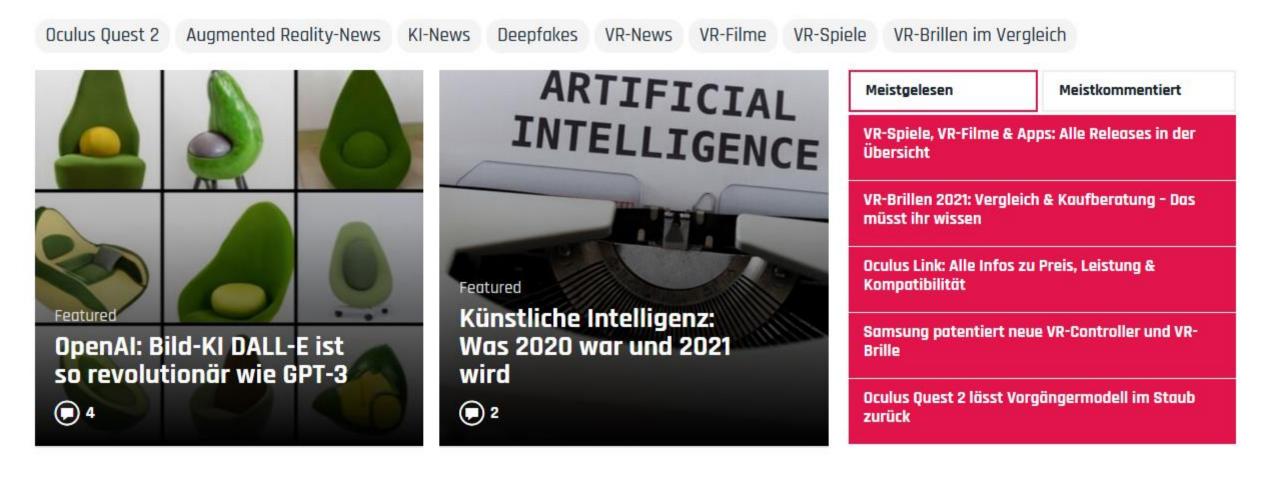
We blocked Facebook from tracking you when the page loaded. If you unblock this page, Facebook will know your activity.

Learn More

Unblock Page

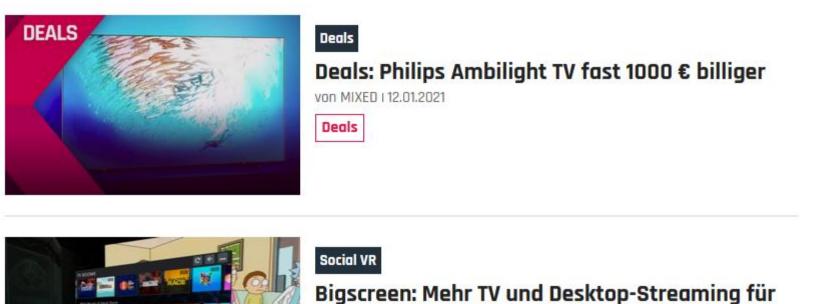






News.

News.



Bigscreen: Mehr TV und Desktop-Streaming für Oculus Quest



Machine Learning



Start writing

Latest Best

Georgia Deaconu in Towards Data Science

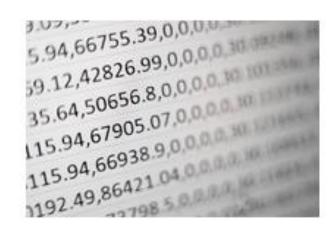


5 ways to deal with large datasets in Python



As a data scientist, I find myself more and more having to deal with "big data". What I abusively call big data...

2 days ago . 3 min read





Q 2



Andrew Lombarti in Towards Data Science







Mistakes I've made on my journey and how you can avoid being like me when starting out

19 hours ago . 5 min read



70K 177K Stories Writers



Related Topics

Artificial Intelligence Data Science Deep Learning Python Technology Programming Neural Networks NLP

Top Writers



Terence Shin







Susan Li

Changing the world, one post at a time. Sr Data Scientist, Toront...







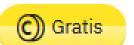
















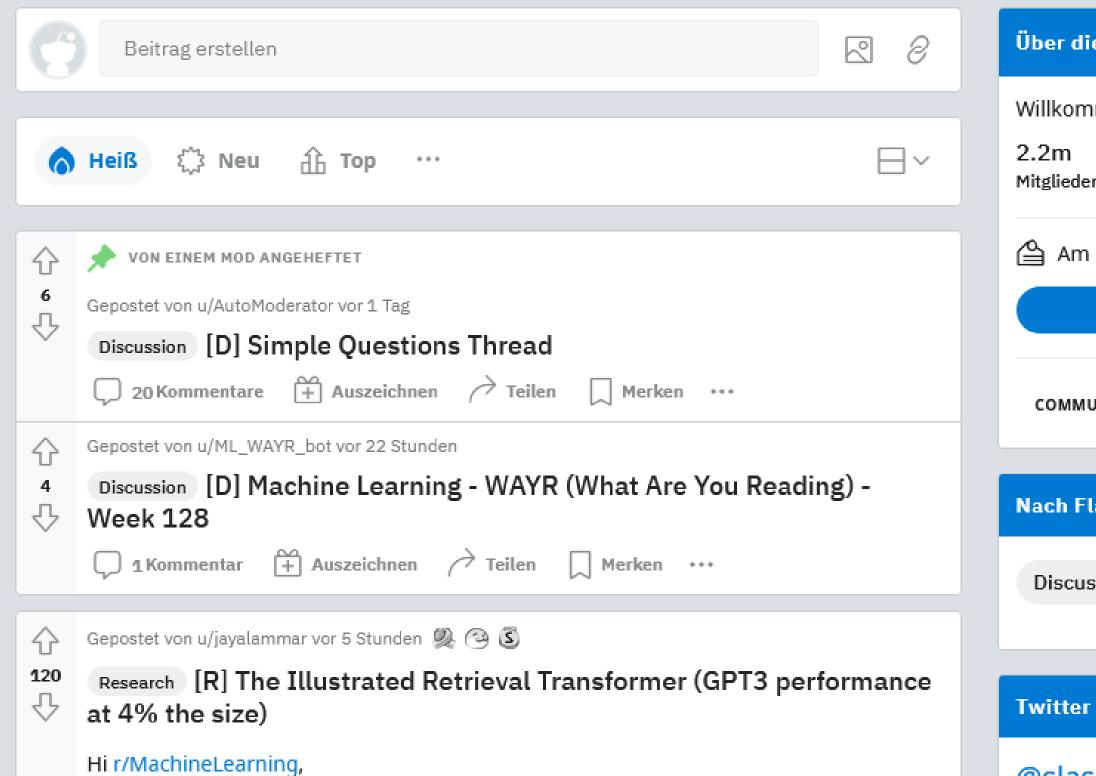
Machine Learning

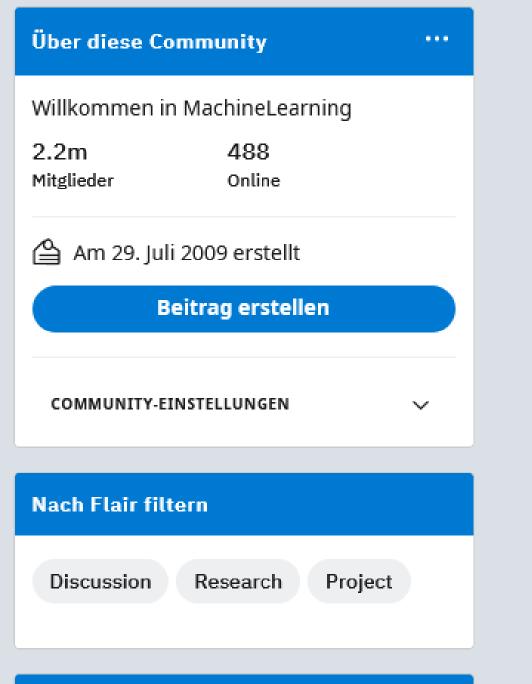
Beigetreten 🔔



r/MachineLearning

Beiträge





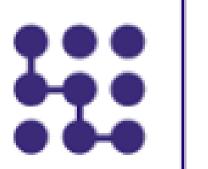


courserd (UDACITY

100 Udemy







:: KI-Campus

Die Lernplattform für Künstliche Intelligenz

OPENCAMPUS.SH MACHINE LEARNING DEGREE

Machine Learning with TensorFlow or Deep Learning