

Vergleich von ARIMA, LSTM und XGBoost zur Modellierung des Stromverbrauchs einer industriellen Anlage

Bachelorarbeit von
Chin-I Feng

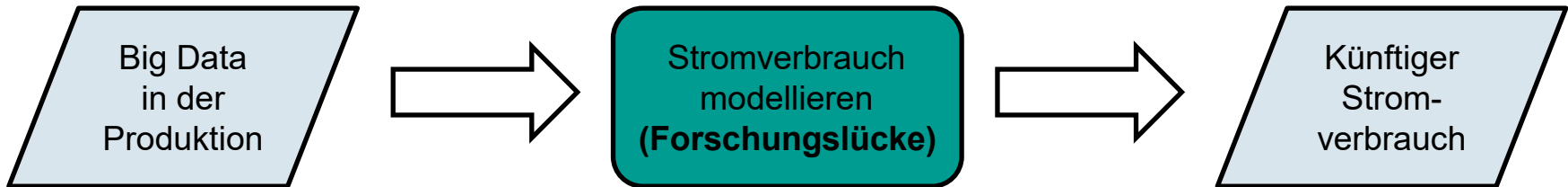
HTW WG I **Agenda**

- Einführung
- Datenverarbeitung
- Feature-Extraktion
- Angewendete Modelle
- Ergebnis
- Anwendung

Einführung

Problemstellung

- Herausforderungen der Datenanalyse bei **wachsenden Datenmengen**.
- Fehlen eines spezifischen Modells für die Vorhersage des Stromverbrauchs in der **Eisengussindustrie**.
- Mangelnde Fähigkeit zur Vorhersage des **Stromverbrauchs**.



Einführung

Zielsetzung

- Entwicklung präziser **Vorhersagemodelle** für den Stromverbrauch einer **Formanlage** der Firma Fondium (Hersteller für Eisengussteile).



Quelle: <https://fondium.eu/>

Einführung

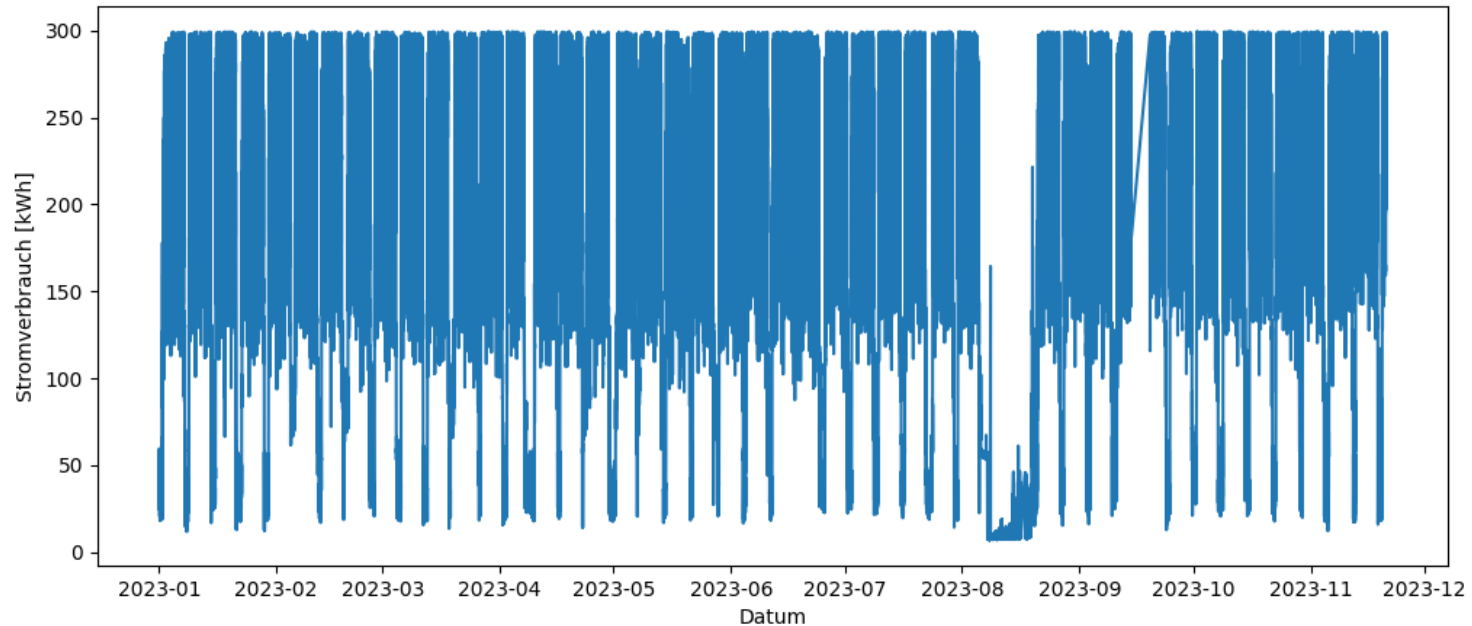
Forschungsfragen

- Wie können Produktionsdaten effektiv als Modell-Eingabe ausgewählt werden, um die Leistung der Vorhersagemodelle zu verbessern?
- Welches Modell bietet die maximale Vorhersagegenauigkeit bei minimaler Trainingsdauer?
- Inwiefern beeinflusst die Optimierung des Stromverbrauchs durch präzise Modellierung und Prognose die Stromkosten?

Datenverarbeitung

Visualisierung

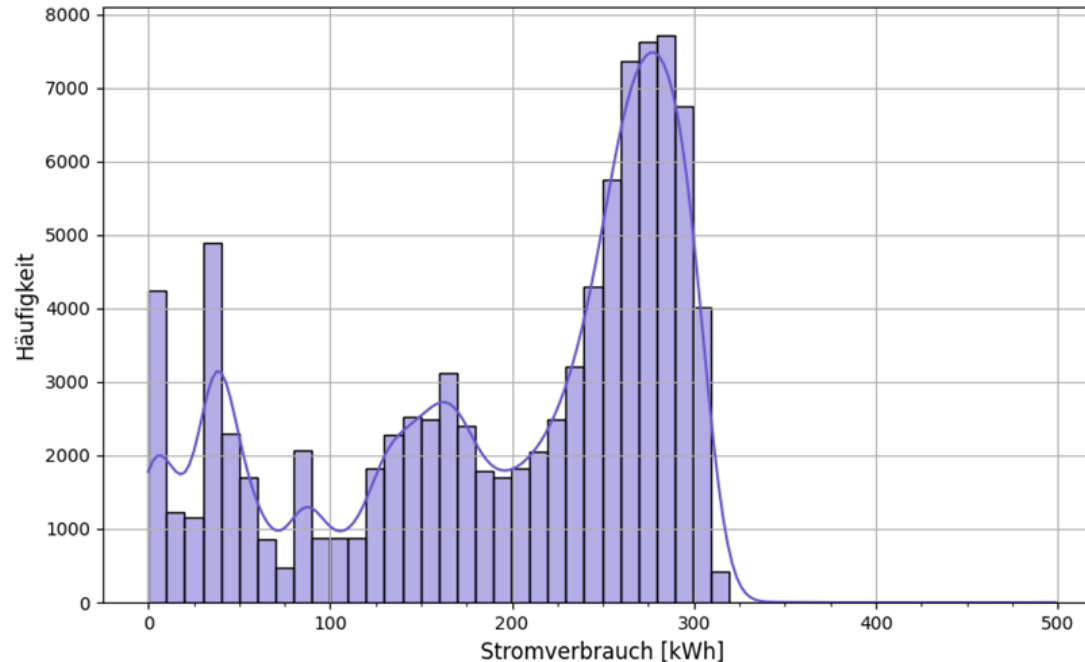
Stromverbrauchsdaten der Formanlage, im 5-Minuten-Takt erfasst vom 01. Januar bis 20. November 2023:



Datenverarbeitung

Bereinigung

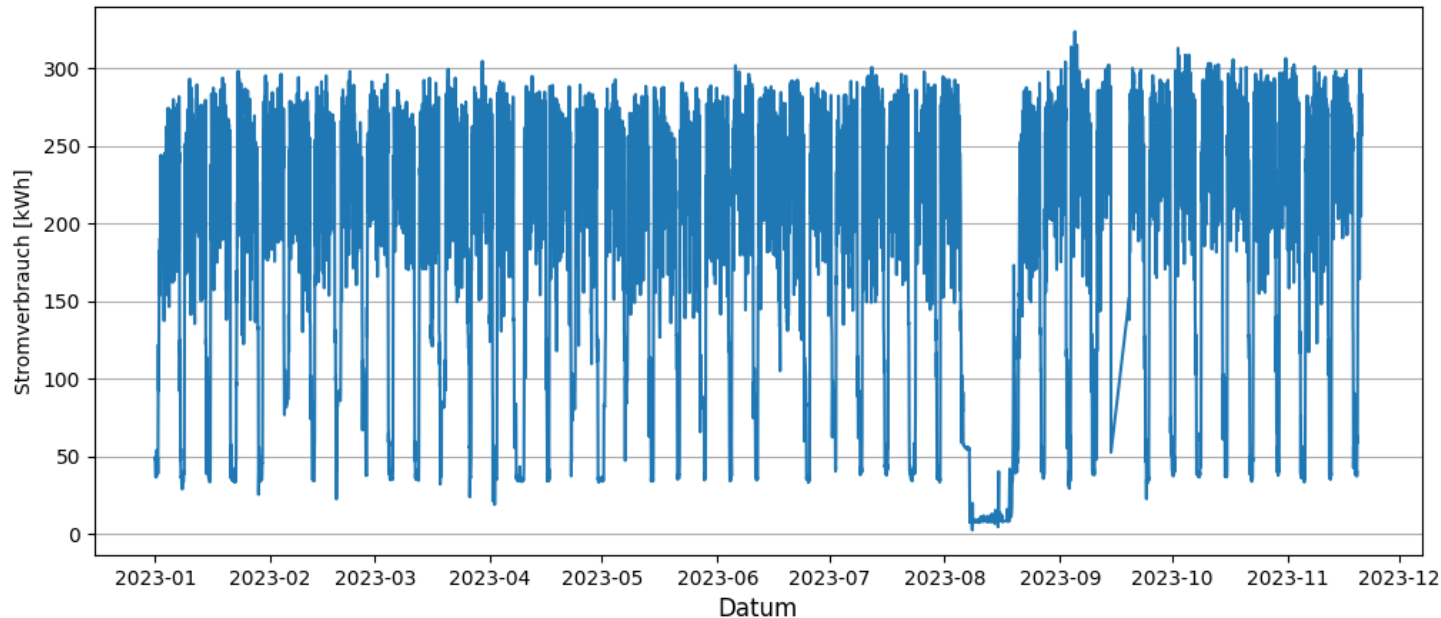
Entfernung von **Ausreißern** zur Gewährleistung der Datenqualität:



Datenverarbeitung

Anpassung der Intervalle

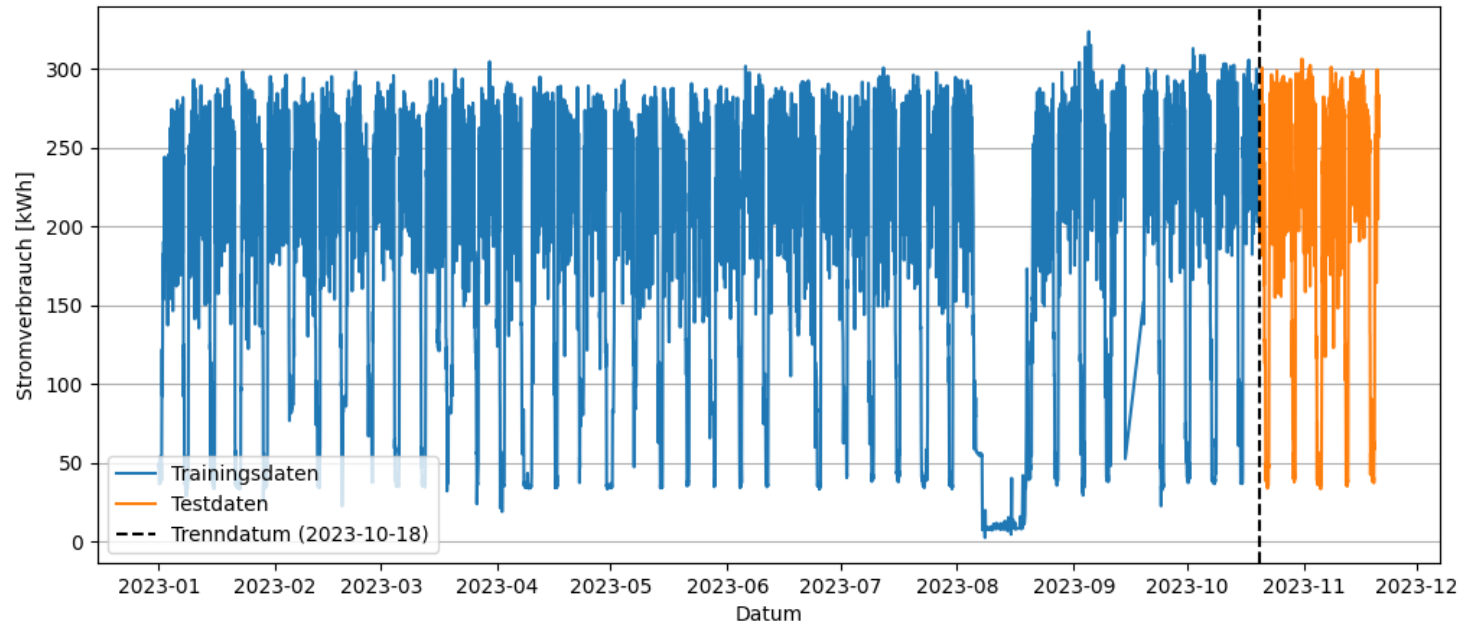
Ursprüngliche Daten, die in 5-Minuten-Intervallen gesammelt wurden, wurden auf **stündliche Intervalle** umgestellt:



Datenverarbeitung

Training und Testaufteilung

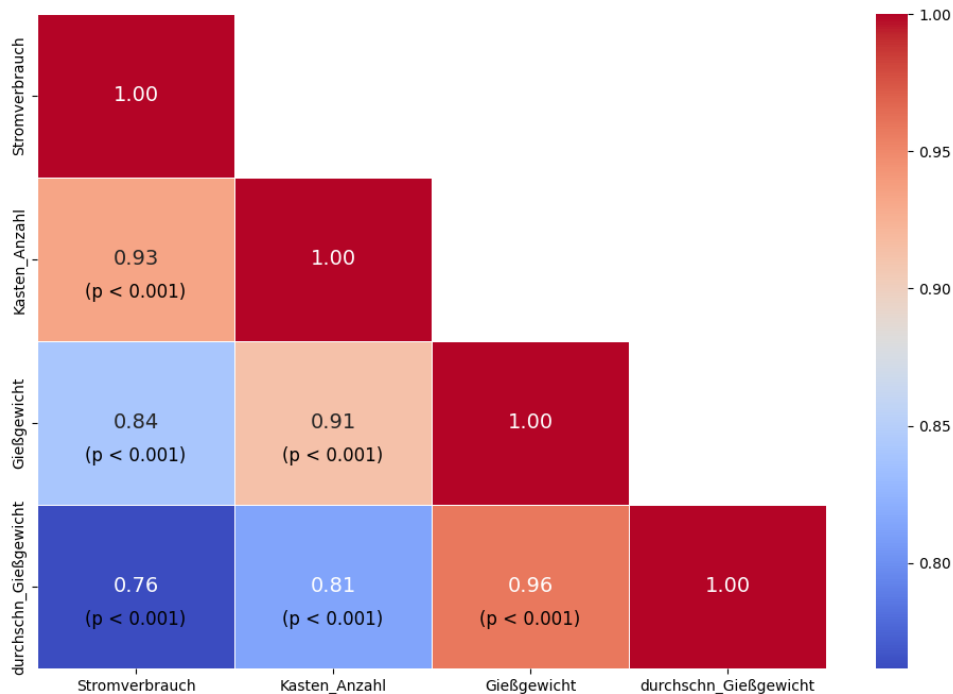
- **Trainingsdaten (90%):** Entwicklung und Validierung des Modells mit Cross Validation
- **Testdaten (10%):** Bewertung der Vorhersagegenauigkeit



Feature-Extraktion

Aussagekräftige Produktionsparameter

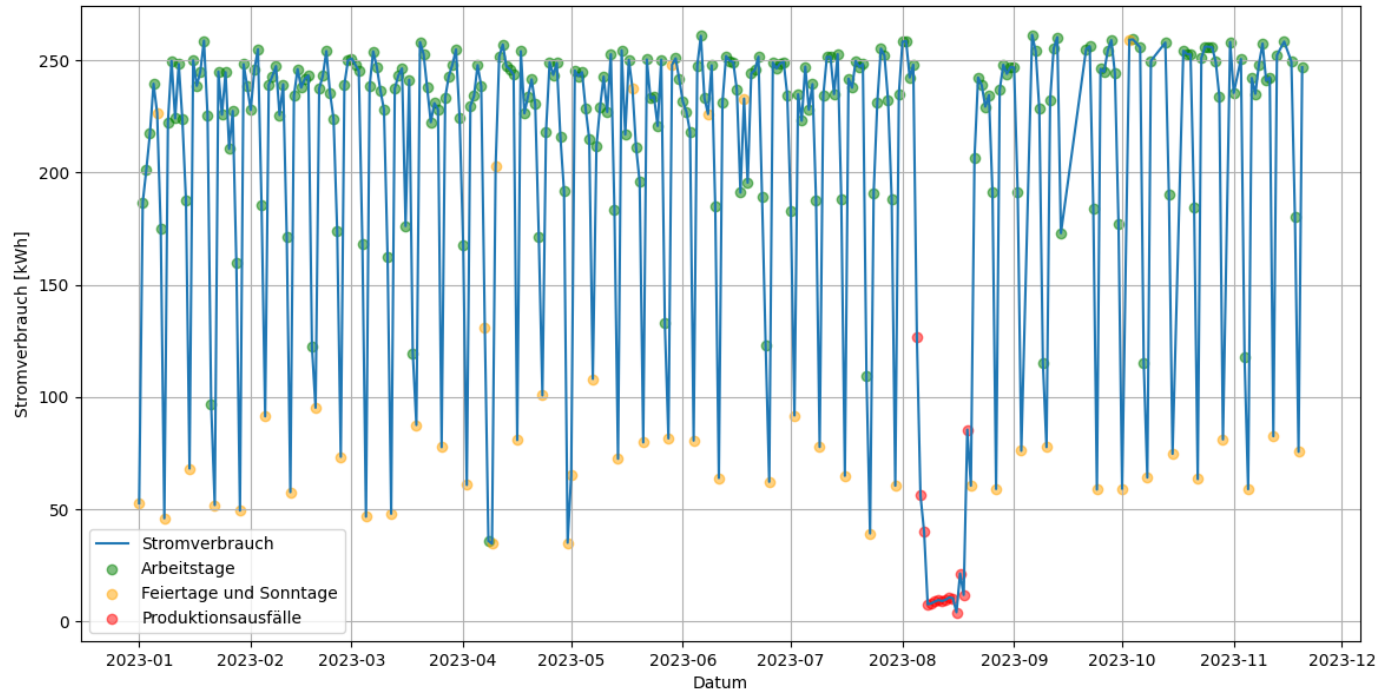
Identifikation von Schlüsselparametern aus den Produktionsdaten, die einen **signifikanten Einfluss** auf den Stromverbrauch haben:



Feature-Extraktion

Tägliche Zeitliche Elemente

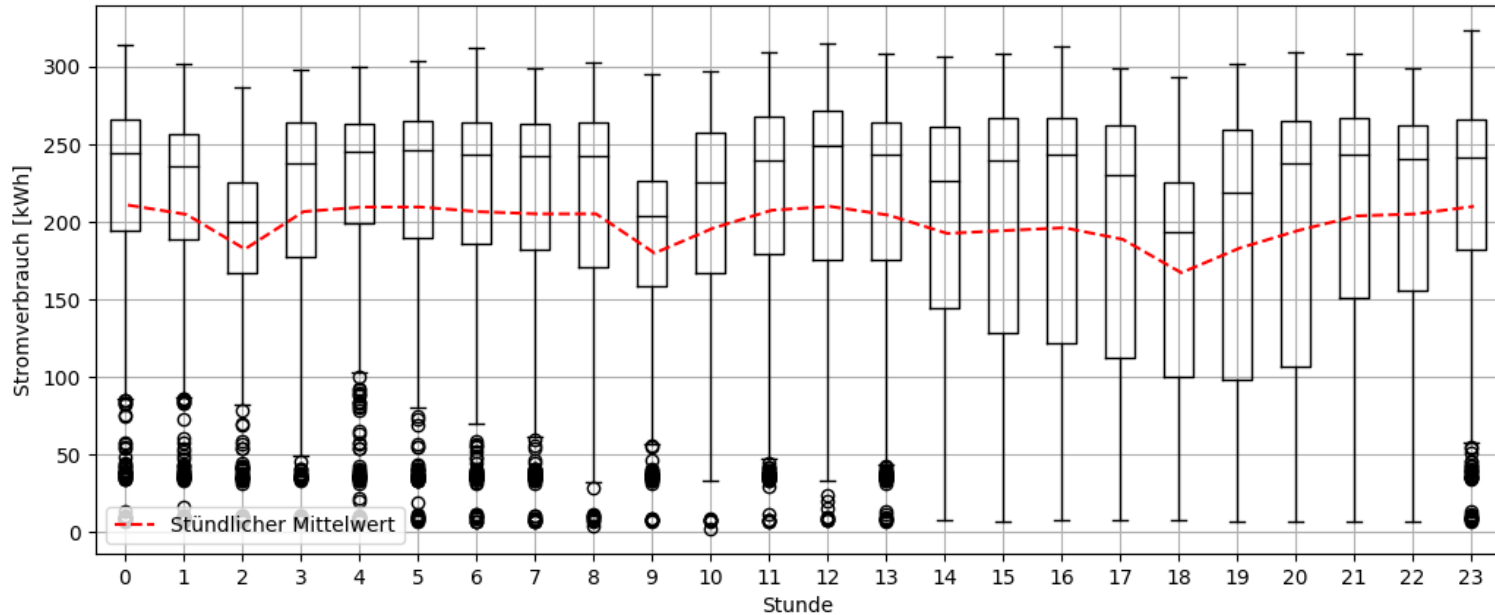
Erfassung und Analyse der täglichen Muster im Stromverbrauch:



Feature-Extraktion

Stündliche Zeitliche Elemente

Untersuchung des stündlichen Stromverbrauchs:



Feature-Extraktion

Eingabeparameter und Zielwert

- **Exogene Variablen (X):** Kastenanzahl, Arbeitstag, Produktionsausfall und Stunde
- **Endogene Variable (y):** Stromverbrauch

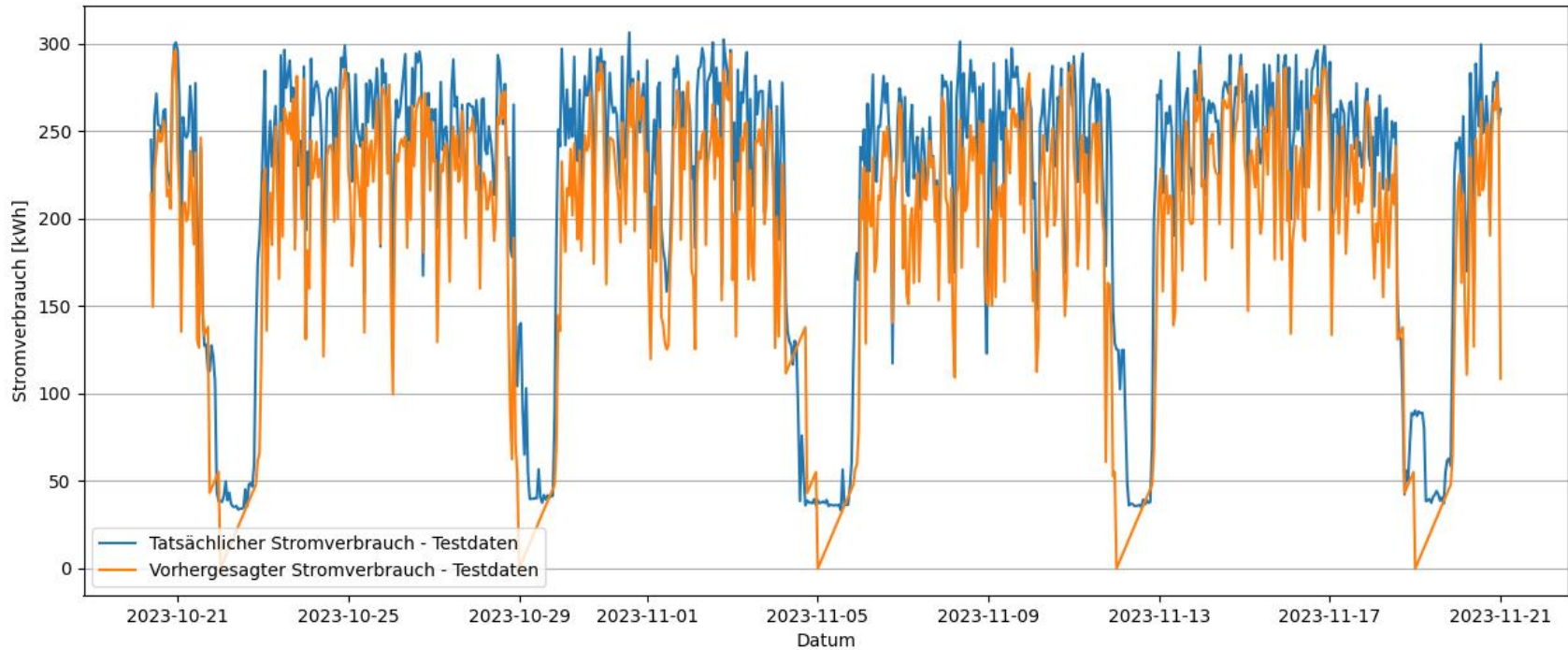


Angewendete Modelle

Modellart	Modell
Statistik	ARIMA (Autoregressive Integrated Moving Average)
Deep-Learning	LSTM (Long Short-Term Memory)
Machine-Learning	XGBoost (Extreme Gradient Boosting)

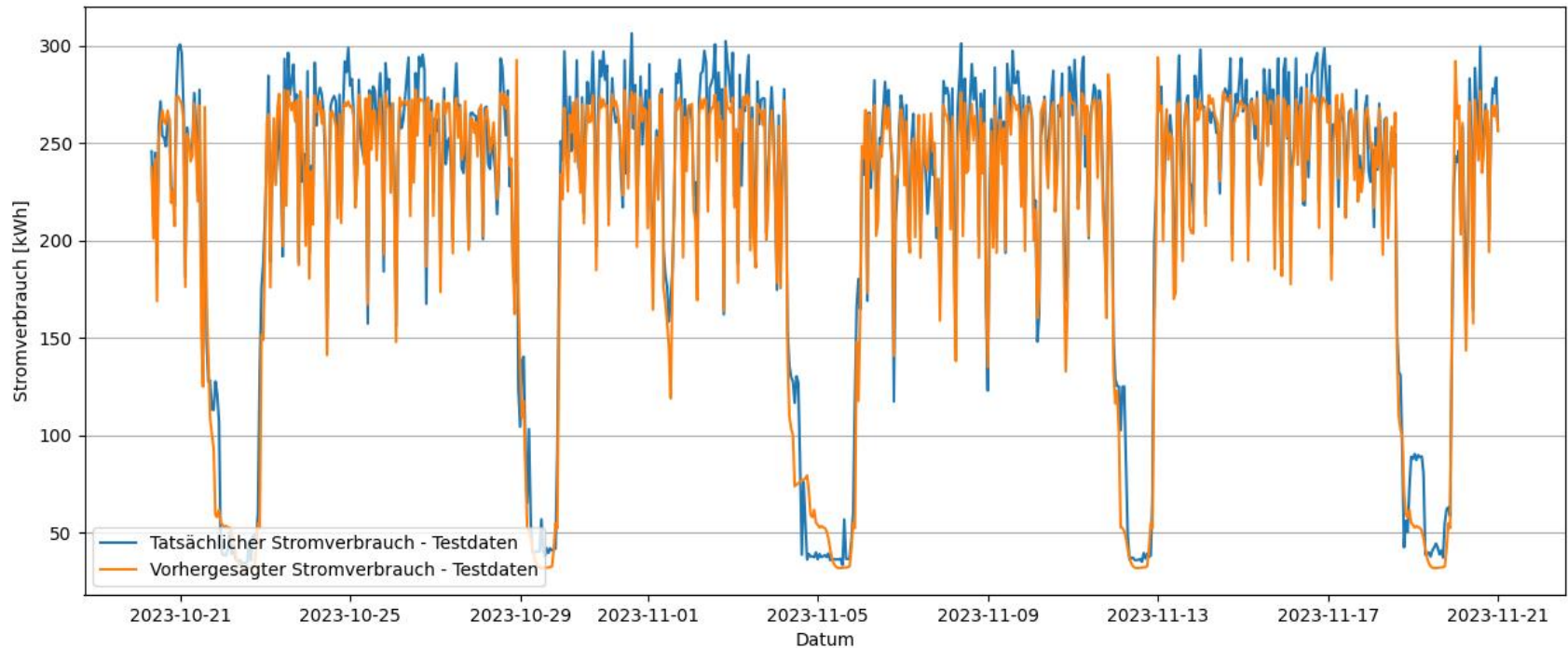
Angewendete Modelle

ARIMA-Vorhersage



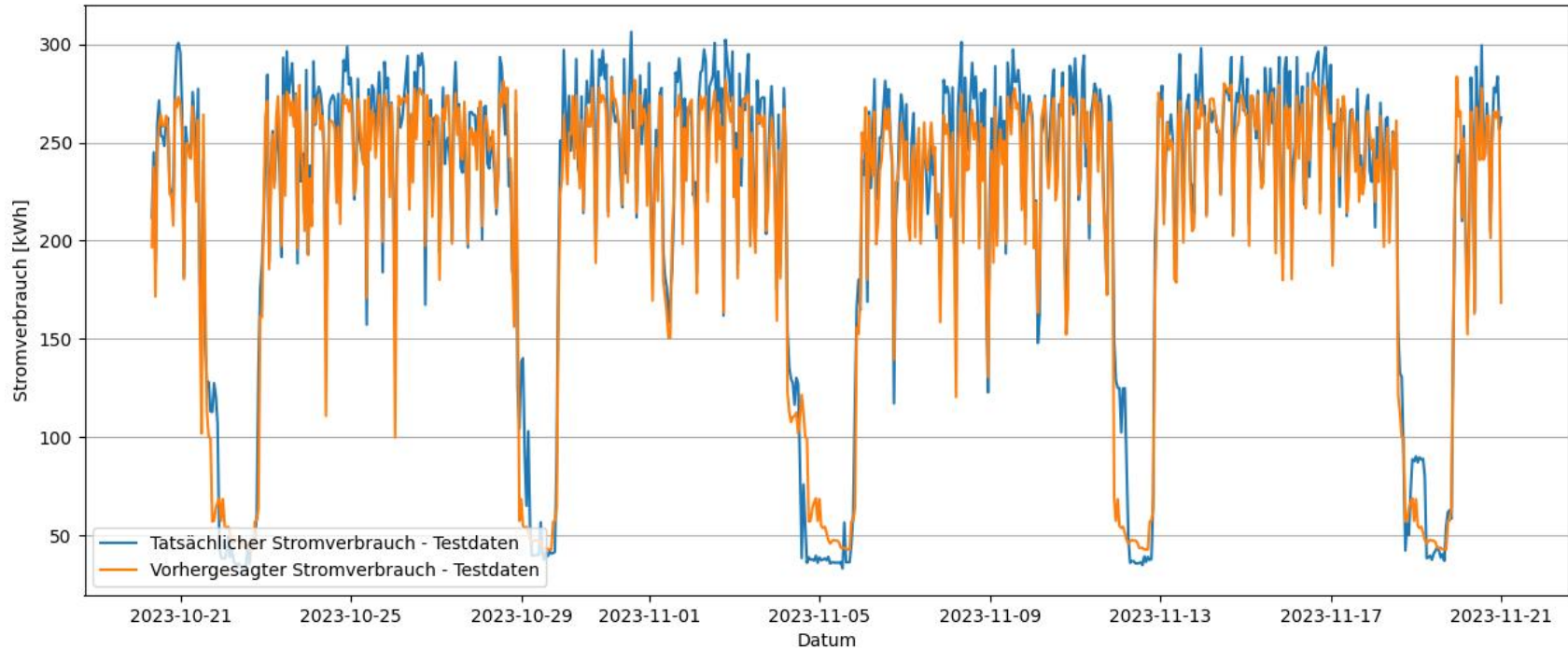
Angewendete Modelle

LSTM-Vorhersage



Angewendete Modelle

XGBoost-Vorhersage



Ergebnis

Leistungsvergleich

Modelle	Validierungsdaten			Testdaten		
	MAE	RMSE	R^2	MAE	RMSE	R^2
ARIMA	33,44	42,62	0,69	32,93	41,14	0,74
LSTM	18,07	23,42	0,91	12,48	16,35	0,95
XGBoost	16,80	21,65	0,92	13,43	18,03	0,95

Ergebnis

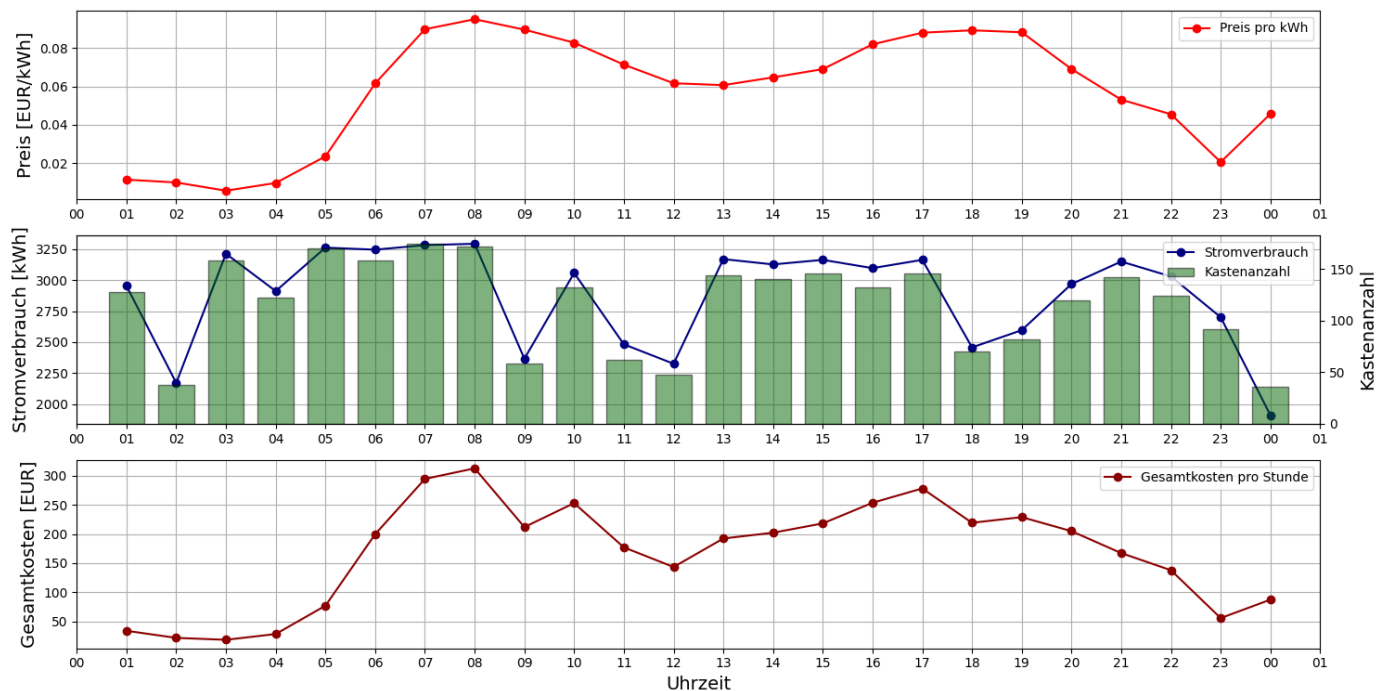
Effizienzvergleich

Modelle	Vorhersage- genauigkeit	Trainings- dauer (s)	Anmerkung
ARIMA	Niedrig	69,679	Nicht geeignet für nicht-lineare Schwankungen und schnelle Änderungen
LSTM	Hoch	35,371	Geeignet für Analysen mit langfristigen Abhängigkeiten aber Einschränkung bei Trainingszeit und Rechenkapazität
XGBoost	Hoch	0,520	Geeignet für komplexe nicht-lineare Beziehungen und schnelle Entscheidungen mit Echtzeitdaten

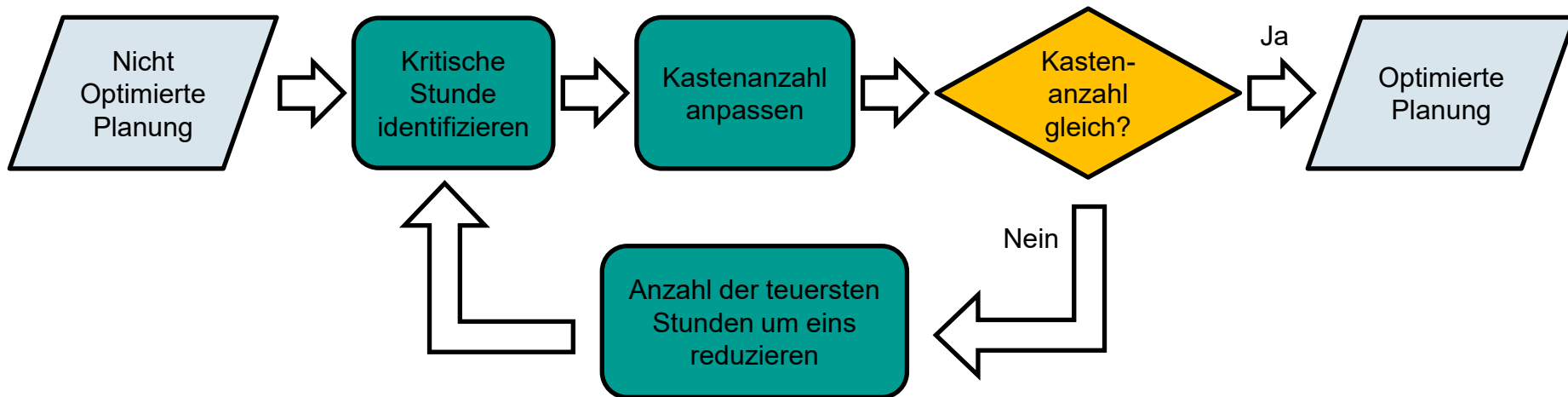
Anwendung

Berechnung der Stromkosten

Formel: Stromkosten = Stromverbrauch (kWh) × Strompreis (€/kWh)



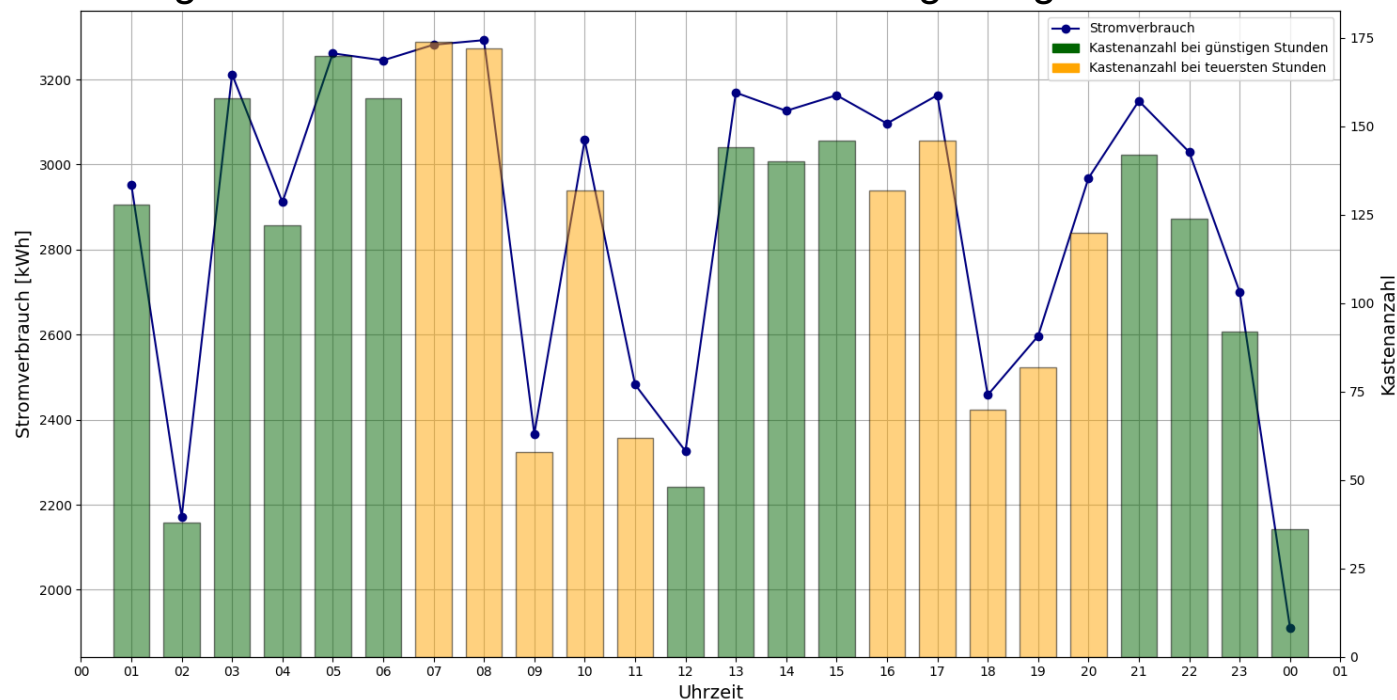
Anpassung der Kastenanzahl (Produktionsplanung) während teurer und günstiger Stromstunden.



Anwendung

Optimierung der Stromkosten

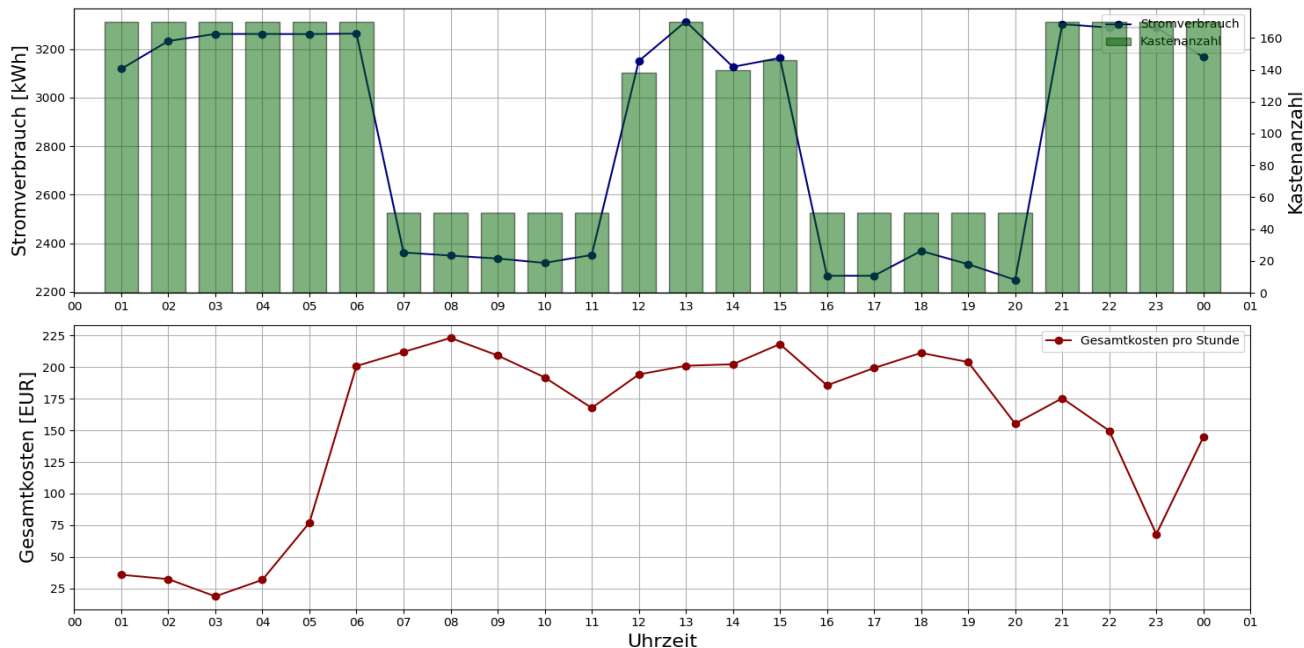
Ermittlung der 10 teuersten Stunden und der 14 günstigsten Stunden:



Anwendung

Optimierung der Stromkosten

Reduzierung der Kastenanzahl in teuren Stunden auf 50 und Erhöhung in günstigen Stunden bis maximal 170:



Urspr. Kosten:
4019,52 €

Optimierte Kosten:
3710,55 €

Einsparpotential:
→ 308,97 €

**Vielen Dank
für die Aufmerksamkeit**

Bildquellen

— <https://fondium.eu/>

Modellvalidierung

Evaluierungsmetriken

- **MAE:** Mittlere absolute Abweichung von tatsächlichen Werten.
- **RMSE:** Quadratwurzel der durchschnittlichen quadrierten Abweichungen.
- **R^2 :** Wie gut das Modell die beobachteten Daten erklärt.
- **Trainingsdauer:** Lernzeit des Modells auf Trainingsdaten.

- Autoregressive Integrated Moving Average
- Ein **stochastischer** Modellierungsansatz, der verwendet werden kann, um die Wahrscheinlichkeit zu berechnen, dass ein zukünftiger Wert zwischen zwei festgelegten Grenzen liegt.

I(d)-Modell

$$z_t^* = \nabla^d z_t$$

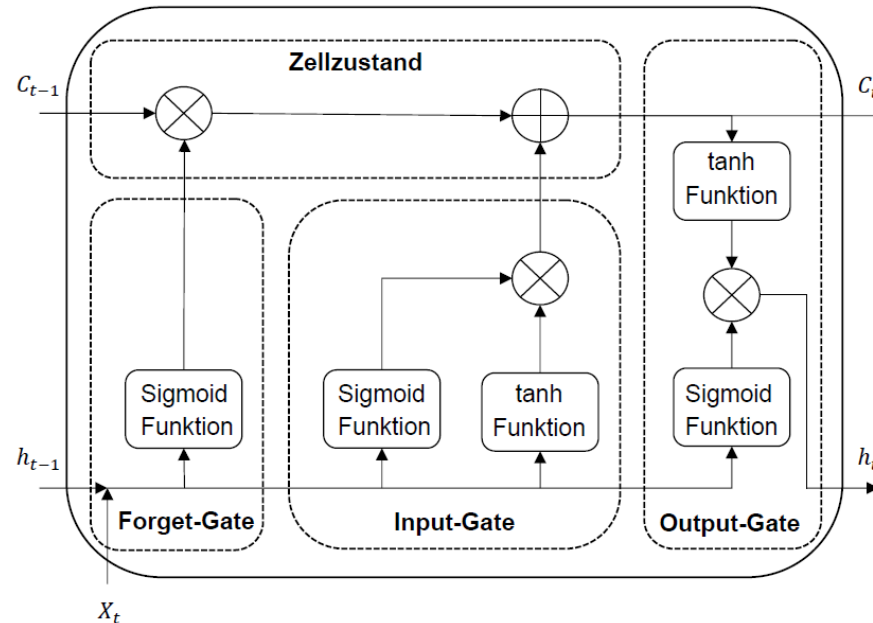
$$z_t^* = \underbrace{\sum_{i=1}^p \phi_i z_{t-i}^* + \epsilon_t}_{\text{AR}(p)\text{-Modell}} - \underbrace{\sum_{i=1}^q \theta_i \epsilon_{t-i}}_{\text{MA}(q)\text{-Modell}}$$

AR(p)-Modell

MA(q)-Modell

LSTM-Modell

- Long Short-Term Memory
- Fortgeschrittene rekurrente neuronale Netze (RNN), die speziell für Deep Learning-Anwendungen in Zeitreihendaten entwickelt wurden.



- eXtreme Gradient Boosted trees
- Ein Machine Learning-Algorithmus, eine Ensemble-Methode, die viele schwache Vorhersagemodelle (Entscheidungsbäume) kombiniert, um robuste Vorhersagen zu erzielen.

