

VORHERSAGE DER LEISTUNG EINER PHOTOVOLTAIKANALGE

Machine Learning

Leon Rottler – BAM – WS 22/23

Agenda

1. Datenbeschaffung
2. Daten Analyse
3. Regressions-Algorithmen
4. Dashboard – Live Demo
5. Code

Problemstellung



Wie gut lässt sich die Leistung
[kWh] einer Photovoltaikanlage
basierend auf Wetterdaten
vorhersagen?

Quelle: https://photovoltaik-rheinland.de/wp-content/uploads/2019/12/Fotolia_haus_solarpanels_baum-1280x854.jpg

Datenbeschaffung

- Zwei Datenquellen
 - Eigene Photovoltaikanlage
 - Wetterdaten des Deutschen Wetterdienstes

Photovoltaikanlage (1/2)



1. Wechselrichter der PV-Anlage per LAN-Kabel ins Netzwerk einbinden

2. Dem Wechselrichter der PV-Anlage eine Feste IPv4-Adresse geben



Solutronic GmbH Webserver für SOLPLUS(Grundmenü)

Navigation Grundmenü Ein-/Ausgang Wechselrichter Sicherheit ENS Umwelt Kommunikation Optionen Datenlogger Datensicherung

SOLPLUS 50, S/N 23720, FW-Version 2.55

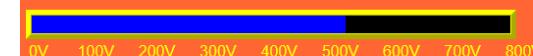
Leistung AC: 636 Watt



Netzspannung: 231 Volt



Gleichspannung: 511 Volt



Einspeisebetrieb

- Netzstrom: 2.80 A
- Strom DC: 1.32 A
- Energie Tag: 1.819 kWh
- Energie Gesamt: 66143 kWh
- Euro-Betrag Heute: 0.516 €
- Euro-Betrag Gesamt: 18784 €
- Wirkungsgrad: 94.0 %
- Temperatur PV-Modul extern: 0 °C
- Einstrahlungsführer extern: 0 W/m²
- Maximale Leistung Heute: 1986 Watt
- Betriebsstunden Heute: 5.6
- Betriebsstunden Gesamt: 50237.0

Anlagendaten

3. Webserver des Wechselrichters ansteuern und txt-Datei herunterladen

Photovoltaikanlage (2/2)

Datei 7 = Energie-Jahres-Logger

P 97	Echtzeit Datum	= 5.12.2022
P 98	Echtzeit Zeit	= 14:06:08
P149	Firmwareversion	= 2.55
P148	Seriennummer	= 23720

Datum Energie-Tag in kWh

01.01. 0013,8

02.01. 0003,1

03.01. 0003,5

04.01. 0000,0

05.01. 0006,6

06.01. 0004,1

07.01. 0001,1

08.01. 0003,4

09.01. 0003,7

10.01. 0010,0

11.01. 0005,9

12.01. 0016,0

13.01. 0014,3

14.01. 0006,7

15.01. 0001,0

16.01. 0008,2

17.01. 0004,1

18.01. 0009,0

19.01. 0011,3

20.01. 0002,1

21.01. 0007,6

22.01. 0002,6

23.01. 0003,3

24.01. 0018,1

25.01. 0017,4

26.01. 0000,7

27.01. 0002,7

Txt-Datei:

- Verschiedene Informationen am Anfang des Dokuments
- Die Leistung der PV-Anlage in kWh pro Tag in zeilenweiser Auflistung darunter
 - Werte werden jährlich überschrieben -> immer genau die Werte des letzten Jahres gespeichert

Daten zum Wetter vom Deutschen Wetterdienst

- Stellt über ein Webportal für verschiedene Messstationen Wetterdaten seit 1950 zur Verfügung (<https://www.dwd.de/DE/leistungen/klimadatendeutschland/klarchivtagmonat.html>)
- Bei download eine txt-Datei, Dateiendung kann aber einfach zu csv geändert und als solche problemlos eingelesen werden

Beispielsweise:

- Tagesmittel der Windgeschwindigkeit
- Tgl. Niederschlagshöhe
- Sonnenscheindauer in h pro Tag
- Schneehöhe Tageswert
- ...

STATIONS_ID	MESS_DATUM	QN_3	FX	FM;QN_4	RSK;RSKF	SDK;SHK_TAG	NM	VPM	PM	TMK	UPM	TXK	TNK	TGK	eor
2812;20210628;	10;	9.4;	2.4;	3; 3.7;	6;	8.767;-999;	7.3;	19.7;	996.30;	22.0;	76.00;	28.1;	17.8;	15.6;eor	
2812;20210629;	10;	14.1;	2.3;	3; 21.0;	6;	7.200;-999;	6.7;	17.7;	995.11;	18.8;	82.30;	23.3;	15.6;	14.6;eor	
2812;20210630;	10;	11.3;	3.9;	3; 18.4;	6;	0.067;-999;	7.8;	15.0;	996.68;	15.3;	86.50;	19.4;	13.5;	11.9;eor	
2812;20210701;	10;	9.8;	3.0;	3; 0.0;	6;	2.567;-999;	7.0;	14.7;	999.11;	16.0;	81.46;	21.0;	12.7;	10.4;eor	
2812;20210702;	10;	6.2;	1.3;	3; 0.0;	0;	10.300;-999;	4.5;	14.3;	1000.06;	17.8;	72.58;	24.2;	11.0;	8.6;eor	
2812;20210703;	10;	11.2;	2.1;	3; 0.0;	6;	8.733;-999;	7.3;	15.6;	997.38;	19.7;	70.29;	27.3;	12.1;	10.2;eor	
2812;20210704;	10;	12.7;	2.5;	3; 13.7;	6;	1.667;-999;	7.9;	17.8;	994.08;	18.4;	84.50;	24.1;	15.7;	14.2;eor	
2812;20210705;	10;	10.9;	3.6;	3; 0.4;	6;	3.000;-999;	6.8;	17.4;	993.29;	19.7;	76.58;	24.5;	16.0;	14.3;eor	
2812;20210706;	10;	8.1;	2.8;	3; 16.6;	6;	1.233;-999;	6.3;	18.1;	993.30;	17.9;	87.71;	21.7;	14.0;	12.6;eor	
2812;20210707;	10;	5.4;	1.7;	3; 0.0;	6;	9.033;-999;	6.0;	16.6;	998.64;	19.3;	75.71;	25.8;	14.3;	13.0;eor	
2812;20210708;	10;	6.8;	1.8;	3; 1.6;	6;	0.000;-999;	7.8;	17.8;	1001.55;	17.7;	88.04;	19.5;	15.9;	13.5;eor	
2812;20210709;	10;	7.6;	2.6;	3; 0.0;	0;	5.600;-999;	6.1;	16.3;	1003.07;	19.0;	75.96;	23.8;	14.2;	11.9;eor	
2812;20210710;	10;	11.3;	1.8;	3; 1.2;	6;	12.267;-999;	6.1;	15.5;	998.53;	19.6;	71.04;	27.2;	12.1;	9.9;eor	
2812;20210711;	10;	5.7;	2.0;	3; 0.0;	6;	7.217;-999;	5.0;	16.2;	998.40;	19.0;	75.13;	23.9;	14.1;	11.8;eor	
2812;20210712;	10;	5.5;	1.5;	3; 35.5;	6;	4.233;-999;	7.2;	18.4;	992.14;	19.4;	82.13;	25.1;	12.5;	10.8;eor	
2812;20210713;	10;	9.9;	3.0;	3; 16.8;	6;	0.450;-999;	7.4;	17.0;	992.60;	16.6;	90.08;	19.1;	13.5;	11.2;eor	
2812;20210714;	10;	8.9;	3.1;	3; 13.2;	6;	0.000;-999;	7.8;	15.9;	993.89;	16.4;	86.08;	21.1;	12.7;	10.8;eor	
2812;20210715;	10;	9.1;	2.9;	3; 1.5;	6;	1.700;-999;	7.7;	17.5;	994.93;	18.4;	83.42;	22.8;	14.5;	13.3;eor	
2812;20210716;	10;	9.0;	1.8;	3; 21.6;	6;	3.400;-999;	7.0;	18.5;	1000.60;	18.3;	88.75;	23.7;	15.3;	13.4;eor	
2812;20210717;	10;	7.0;	2.7;	3; 0.0;	0;	1.233;-999;	7.0;	19.1;	1003.50;	19.2;	85.92;	23.0;	15.7;	13.6;eor	
2812;20210718;	10;	10.5;	3.6;	3; 0.0;	0;	14.367;-999;	0.0;	18.2;	1001.61;	20.6;	76.25;	25.9;	14.2;	11.2;eor	
2812;20210719;	10;	9.2;	3.3;	3; 0.0;	6;	14.433;-999;	0.0;	18.0;	999.69;	20.8;	75.08;	27.2;	13.9;	11.4;eor	
2812;20210720;	10;	7.8;	2.9;	3; 0.0;	0;	13.983;-999;	0.1;	16.0;	1000.55;	19.0;	73.96;	24.6;	13.4;	10.3;eor	
2812;20210721;	10;	7.5;	2.7;	3; 0.0;	0;	13.883;-999;	0.8;	16.3;	1002.37;	19.2;	74.58;	25.7;	11.9;	9.5;eor	
2812;20210722;	10;	7.6;	1.9;	3; 0.0;	0;	14.067;-999;	0.3;	16.5;	1001.21;	20.1;	71.79;	26.4;	13.3;	10.3;eor	
2812;20210723;	10;	6.0;	1.6;	3; 0.6;	6;	13.683;-999;	2.4;	18.0;	996.26;	21.0;	74.08;	28.1;	13.3;	10.4;eor	
2812;20210724;	10;	7.9;	2.1;	3; 0.4;	6;	4.200;-999;	7.5;	19.7;	991.77;	21.7;	77.67;	27.3;	17.9;	15.6;eor	
2812;20210725;	10;	7.4;	2.3;	3; 0.0;	6;	7.850;-999;	5.0;	19.1;	992.09;	21.4;	76.58;	26.5;	16.7;	14.4;eor	
2812;20210726;	10;	14.6;	2.0;	3; 20.6;	6;	8.417;-999;	5.6;	19.1;	994.53;	20.2;	80.83;	25.1;	16.0;	13.7;eor	

- PV-Anlage Daten Zeile für Zeile ausgelesen
 - Anhand eines Musters die Zeilen mit den kWh-Werten in einen pandas DataFrame umgewandelt
 - Datumsformat angepasst und Datum als Index gesetzt
- Wetter Daten aus der CSV Datei gelesen
 - Irrelevante Spalten entfernt (bspw. Stations_id, eor)
 - Formatierungsfehler behoben (im wesentliche Leerzeichen die zu Problemen geführt haben)
 - Datumsformat angepasst und als Index gesetzt
- Beide Datensätze mittels des Datums zusammen geführt

Daten Analyse

Daten – Deskriptive Statistik

First 10 rows of the data:

	kWh	QN_3	FX	FM	QN_4	...	TMK	UPM	TXK	TNK	TGK
MESS_DATUM						...					
2022-01-01	13.8	10	4.6	1.5	3	...	7.3	87.29	14.0	2.8	0.5
2022-01-02	3.1	10	11.3	3.2	3	...	8.2	78.54	13.9	2.4	1.2
2022-01-03	3.5	10	12.2	5.4	3	...	11.4	73.33	12.7	9.7	8.1
2022-01-04	0.0	10	19.2	6.3	3	...	10.7	77.75	15.7	5.6	4.7
2022-01-05	6.6	10	12.0	4.1	3	...	4.1	72.42	6.2	2.3	0.9
2022-01-06	4.1	10	7.1	2.5	3	...	2.2	87.13	5.4	-2.3	-4.0
2022-01-07	1.1	10	11.6	4.7	3	...	1.7	82.79	5.0	-2.2	-3.1
2022-01-08	3.4	10	14.2	6.2	3	...	4.6	74.04	7.3	1.7	-1.0
2022-01-09	3.7	10	18.7	4.5	3	...	4.5	85.29	7.5	1.0	-0.4
2022-01-10	10.0	10	4.5	1.4	3	...	1.8	93.33	5.3	-1.9	-2.8

[10 rows x 17 columns]

Dimension of the data:

(363, 17)

Description of the data:

	count	mean	std	min	25%	50%	75%	max
Name: kWh, dtype: float64	363.000000	17.272452	10.617254	0.000000	7.400000	18.600000	27.200000	35.200000
Name: QN_3, dtype: float64	363.000000	5.509642	3.863956	1.000000	3.000000	3.000000	10.000000	10.000000

Name: QN_3, dtype: float64

Data types for each attribute:

kWh	float64
QN_3	int64
FX	float64
FM	float64
QN_4	int64
RSK	float64
RSKF	int64
SDK	float64
SHK_TAG	int64
NM	float64
VPM	float64
PM	float64
TMK	float64
UPM	float64
TXK	float64
TNK	float64
TGK	float64
dtype: object	

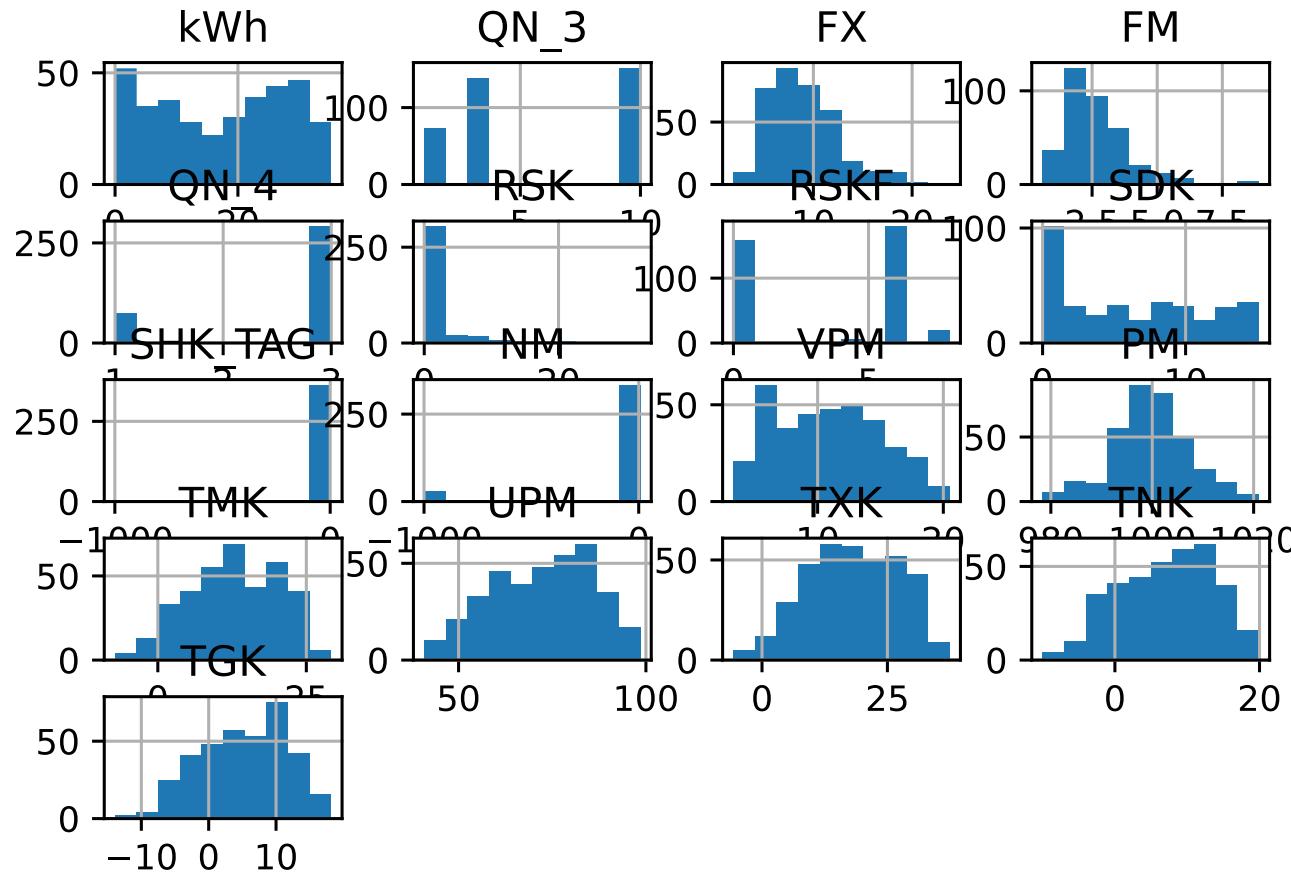
Description of the data:

	kWh	QN_3	FX	...	TXK	TNK	TGK
count	363.000000	363.000000	363.000000	...	363.000000	363.000000	363.000000
mean	17.272452	5.509642	9.006061	...	18.096970	7.048760	5.144628
std	10.617254	3.863956	3.563088	...	9.147256	6.448583	6.445864
min	0.000000	1.000000	1.900000	...	-5.700000	-10.000000	-13.900000
25%	7.400000	3.000000	6.400000	...	11.000000	2.100000	0.150000
50%	18.600000	3.000000	8.500000	...	18.400000	7.700000	5.700000
75%	27.200000	10.000000	11.100000	...	25.800000	12.300000	9.900000
max	35.200000	10.000000	23.800000	...	37.400000	19.900000	18.200000

[8 rows x 17 columns]

Daten - Überblick

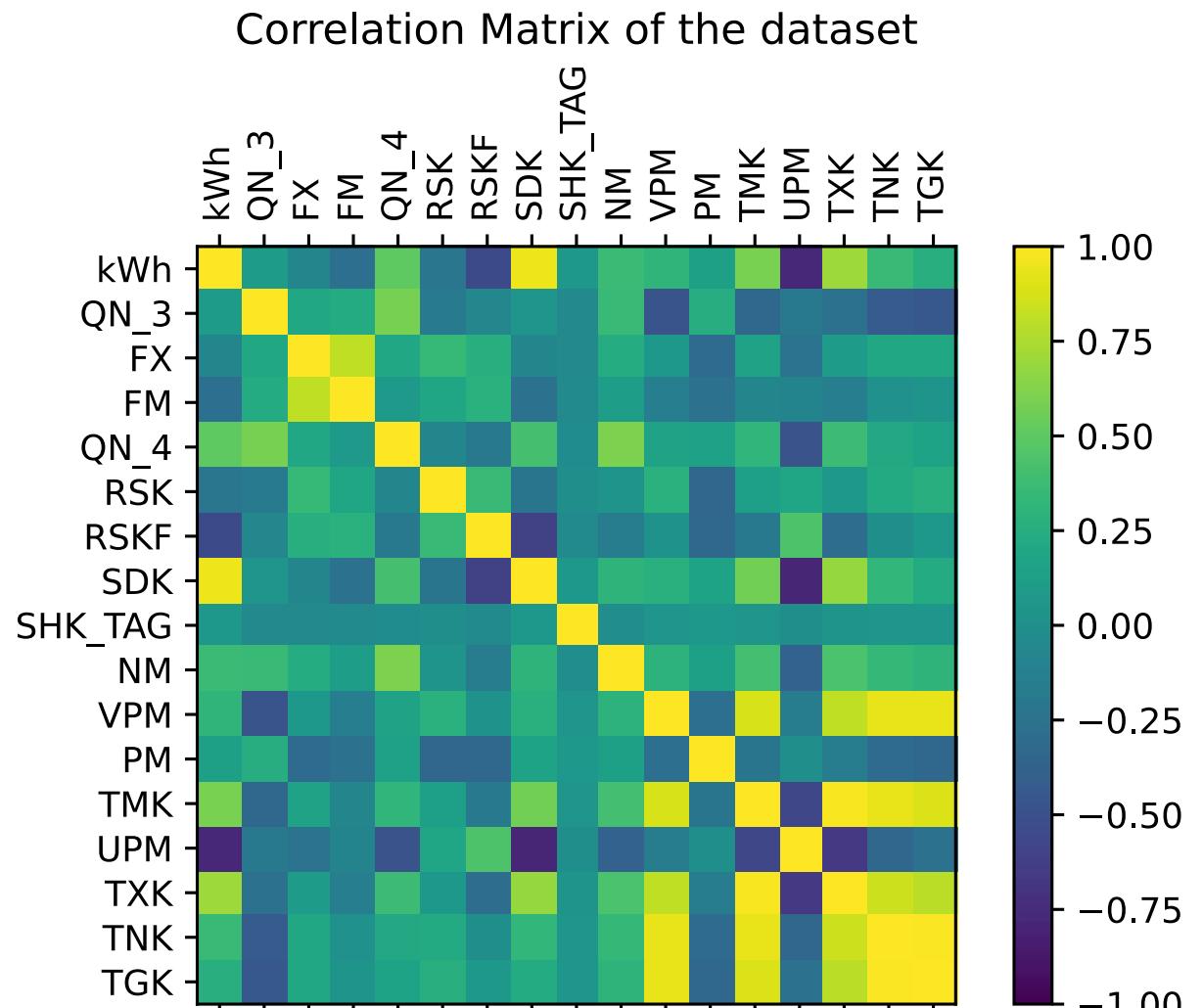
Histogramm aller Attribute:



Legende:

- FX = Maximum der Windspitze
- FM = Tagesmittel Windgeschwind.
- RSK = tgl. Niederschlagshöhe
- RSKF = tgl. Niederschlagsform
- SDK = Sonnenscheindauer (Tag)
- SHK_TAG = Schneehöhe (Tag)
- NM = Tagesmittel des Bedeckungsgrades
- VPM = Tagesmittel des Dampfdruck
- PM = Tagesmittel des Luftdrucks
- TMK = Tagesmittel der Temperatur
- UPM = Tagesmittel der relativen Feuchte
- TXK = Tagesmaximum der Lufttemperatur (2m Höhe)
- TNK = Tagesminimum der Lufttemperatur (2m Höhe)
- TGK = Minimum der Lufttemperatur am Erdboden (5cm Höhe)

Daten – Korrelation der Attribute



Legende:

- FX = Maximum der Windspitze
- FM = Tagesmittel Windgeschwind.
- RSK = tgl. Niederschlagshöhe
- RSKF = tgl. Niederschlagsform
- SDK = Sonnenscheindauer (Tag)
- SHK_TAG = Schneehöhe (Tag)
- NM = Tagesmittel des Bedeckungsgrades
- VPM = Tagesmittel des Dampfdruck
- PM = Tagesmittel des Luftdrucks
- TMK = Tagesmittel der Temperatur
- UPM = Tagesmittel der relativen Feuchte
- TXK = Tagesmaximum der Lufttemperatur (2m Höhe)
- TNK = Tagesminimum der Lufttemperatur (2m Höhe)
- TGK = Minimum der Lufttemperatur am Erdboden (5cm Höhe)

Daten – Überlegungen zu den Attributen

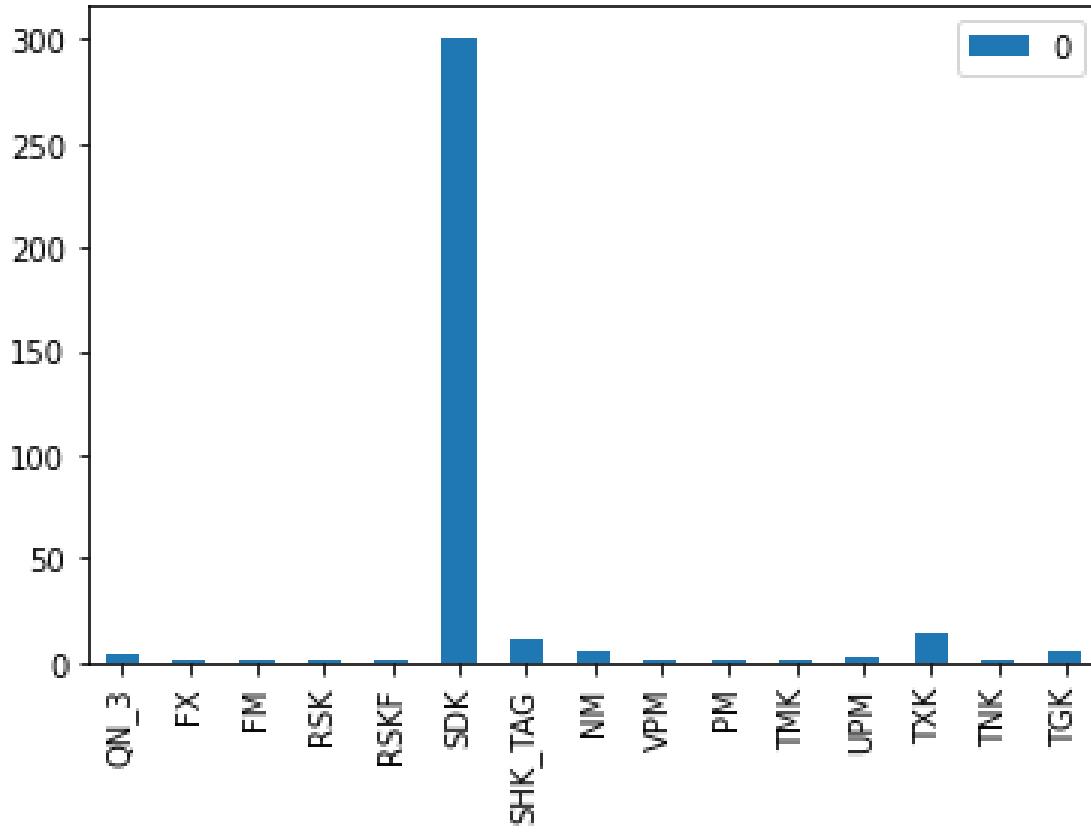
Vermutlich nicht alle Attribute für die Vorhersage relevant:

- Tgl. Niederschlagsform
- Tagesmittel Windgeschwindigkeit

→ Erstes Modell trainieren und die Wichtigkeit der einzelnen Attribute (feature importance) ansehen

Daten – Feature importance

Feature importance für erstes/einfach trainierte XGBoost-Modell:



→ Vorhersage stark Abhängig von Sonnenscheindauer pro Tag

- Es gibt relativ unwichtige Attribute:
 - Maximum der Windspitze
 - Tagesmittel Windgeschwind.
 - Tagesmittel des Dampfdruck
 - ...

→ Diese Attribute werden aus dem Datensatz entfernt

Daten V2 – Deskriptive Statistik

```
First 10 rows of the data:  
      kwh   RSK    SDK  SHK_TAG    NM    TMK    TXK    TNK  
MESS_DATUM  
2022-01-01  13.8    0.0  6.933        0  5.3    7.3  14.0    2.8  
2022-01-02   3.1    0.0  1.583        0  7.2    8.2  13.9    2.4  
2022-01-03   3.5    1.2  0.150        0  8.0   11.4  12.7    9.7  
2022-01-04   0.0   10.7  0.000        0  7.8   10.7  15.7    5.6  
2022-01-05   6.6    1.2  2.317        0  7.2    4.1   6.2    2.3  
2022-01-06   4.1    0.0  0.833        0  5.5    2.2   5.4   -2.3  
2022-01-07   1.1    2.1  0.000        0  7.5    1.7   5.0   -2.2  
2022-01-08   3.4    2.6  0.200        0  7.7    4.6   7.3    1.7  
2022-01-09   3.7    3.4  0.650        0  7.2    4.5   7.5    1.0  
2022-01-10  10.0    0.0  2.850        0  5.5    1.8   5.3   -1.9  
-----  
Dimension of the data:  
(363, 8)
```

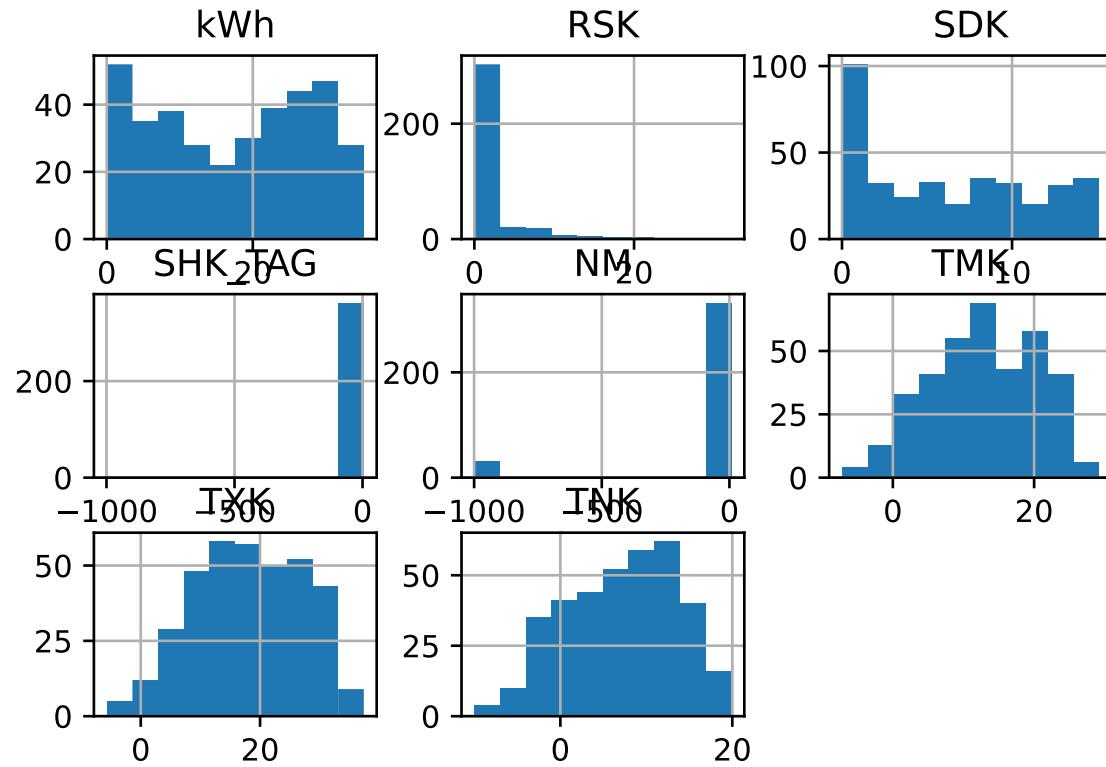
```
Description of the data:  
      kwh       RSK      SDK ...      TMK      TXK      TNK  
count  363.000000  363.000000  363.000000 ...  363.000000  363.000000  363.000000  
mean   17.272452  1.895317  6.200196 ...  12.715978  18.096970  7.048760  
std    10.617254  4.642728  4.988143 ...  7.555637  9.147256  6.448583  
min    0.000000  0.000000  0.000000 ... -7.200000 -5.700000 -10.000000  
25%   7.400000  0.000000  1.133500 ...  7.350000 11.000000  2.100000  
50%  18.600000  0.000000  5.667000 ... 12.300000 18.400000  7.700000  
75%  27.200000  1.100000 10.417000 ... 19.450000 25.800000 12.300000  
max  35.200000 32.200000 15.117000 ... 29.200000 37.400000 19.900000  
[8 rows x 8 columns]
```

```
Description of the data:  
      count      mean      std      min      25%      50%      75%      max  
kwh    363.000000  17.272452  10.617254  0.000000  7.400000  18.600000  27.200000  35.200000  
Name: kwh, dtype: float64  
-----  
      count      mean      std      min      25%      50%      75%      max  
QN_3   363.000000  5.509642  3.863956  1.000000  3.000000  3.000000 10.000000 10.000000  
Name: QN_3, dtype: float64
```

```
Data types for each attribute:  
kwh      float64  
RSK      float64  
SDK      float64  
SHK_TAG  int64  
NM       float64  
TMK      float64  
TXK      float64  
TNK      float64  
dtype: object  
-----  
Data skew:  
kwh     -0.098202  
QN_3    0.230134  
FX      0.956005  
FM      1.780050  
QN_4    -1.497613  
RSK     3.647151  
RSKF    -0.171913  
SDK     0.213740  
SHK_TAG -19.050192  
NM      -2.978907  
VPM     0.187811  
PM      -0.063582  
TMK     -0.179522  
UPM     -0.233394  
TXK     -0.154076  
TNK     -0.279420  
TGK     -0.287010  
dtype: float64
```

Daten V2 - Überblick

Histogramm aller Attribute des neuen Datensatzes:

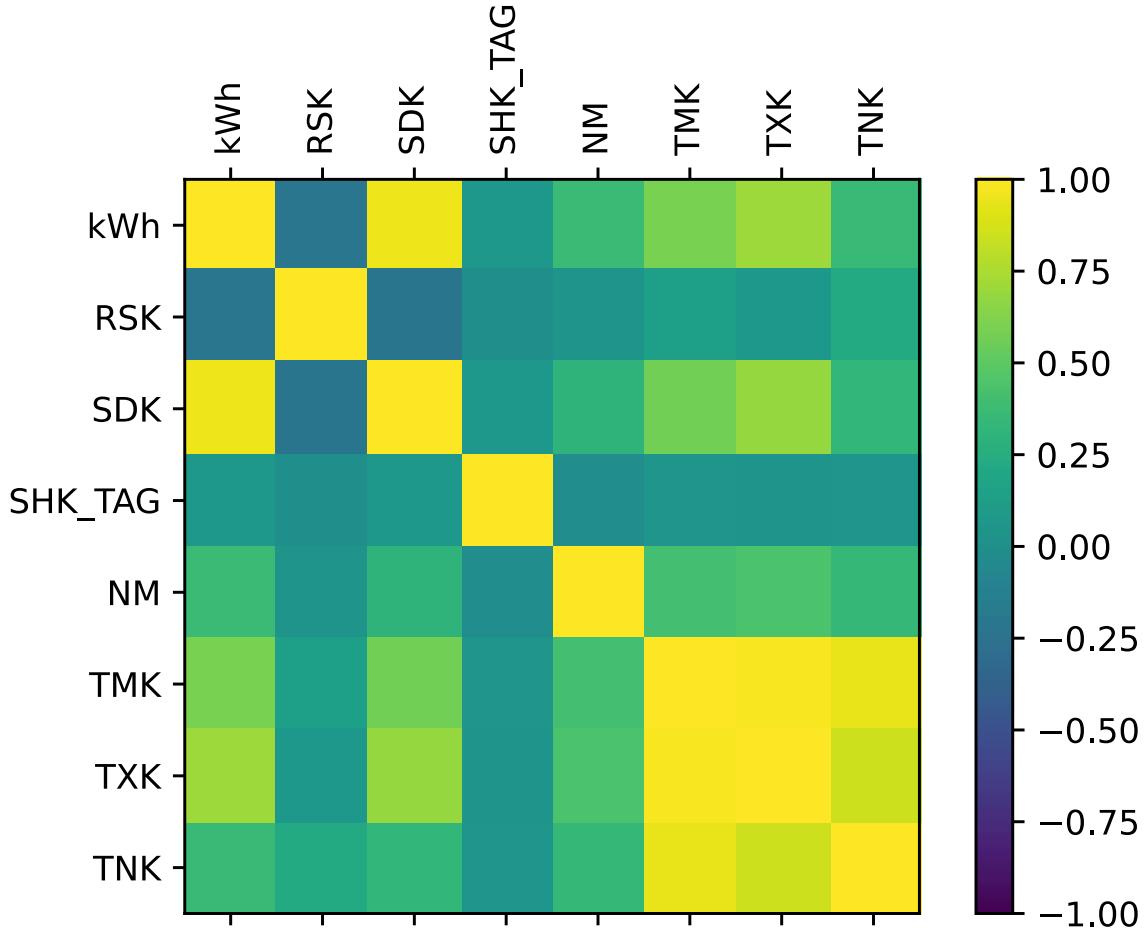


Legende:

- RSK = tgl. Niederschlagshöhe
- SDK = Sonnenscheindauer (Tag)
- SHK_TAG = Schneehöhe (Tag)
- NM = Tagesmittel des Bedeckungsgrades
- TMK = Tagesmittel der Temperatur
- TXK = Tagesmaximum der Lufttemperatur (2m Höhe)
- TNK = Tagesminimum der Lufttemperatur (2m Höhe)

Daten V2 – Korrelation der Attribute

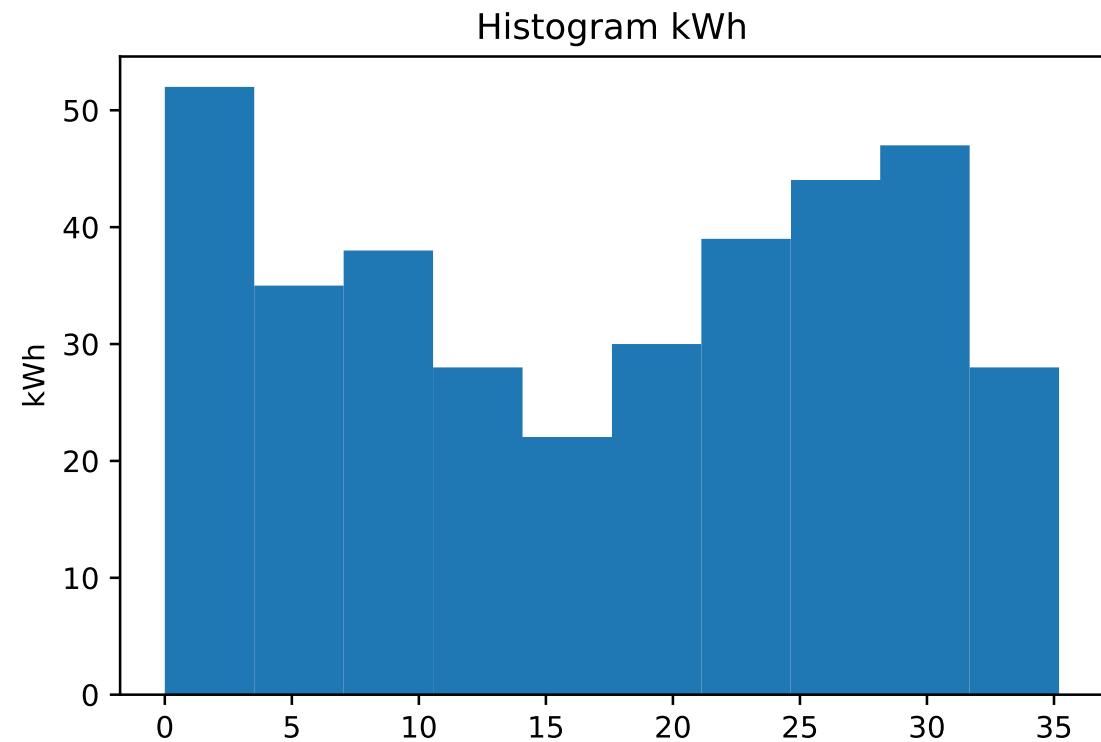
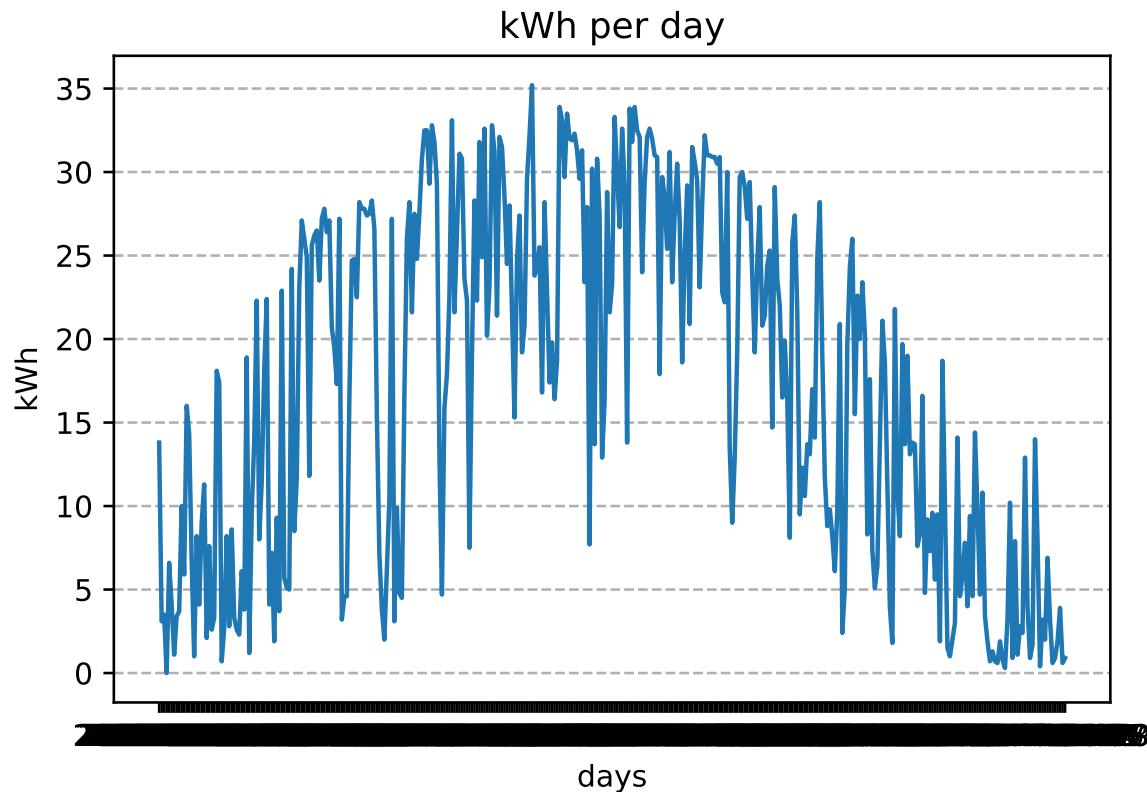
Correlation Matrix of the dataset V2



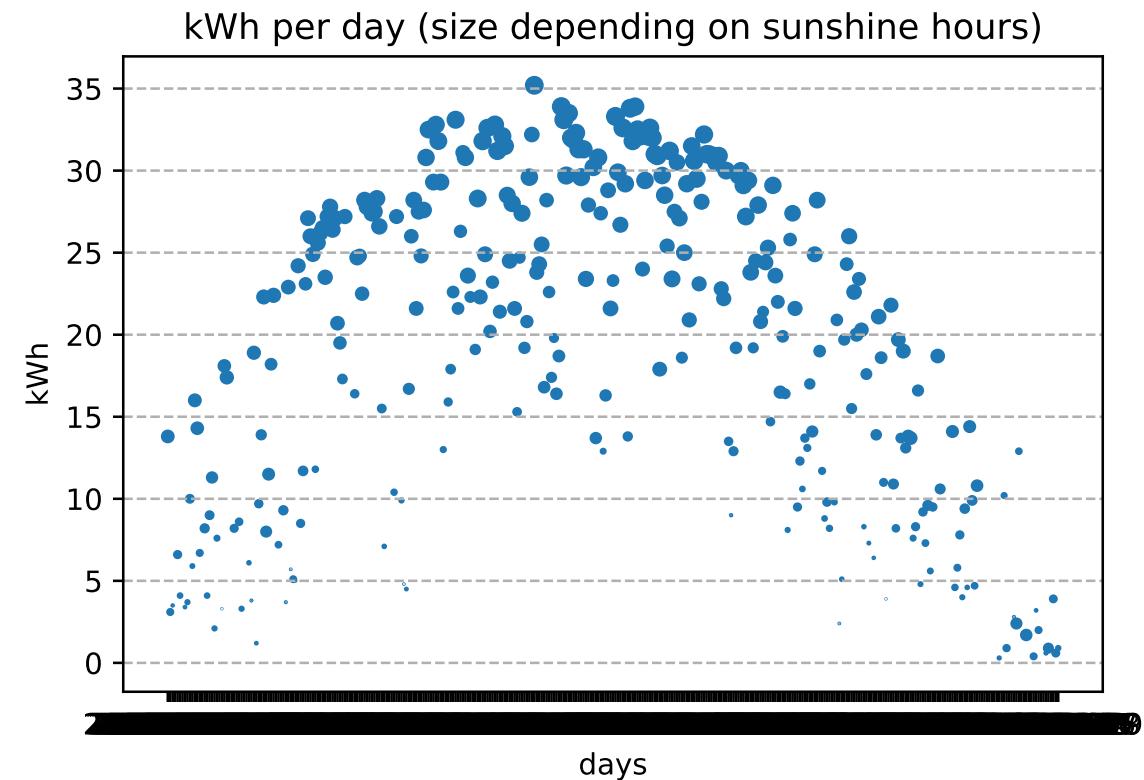
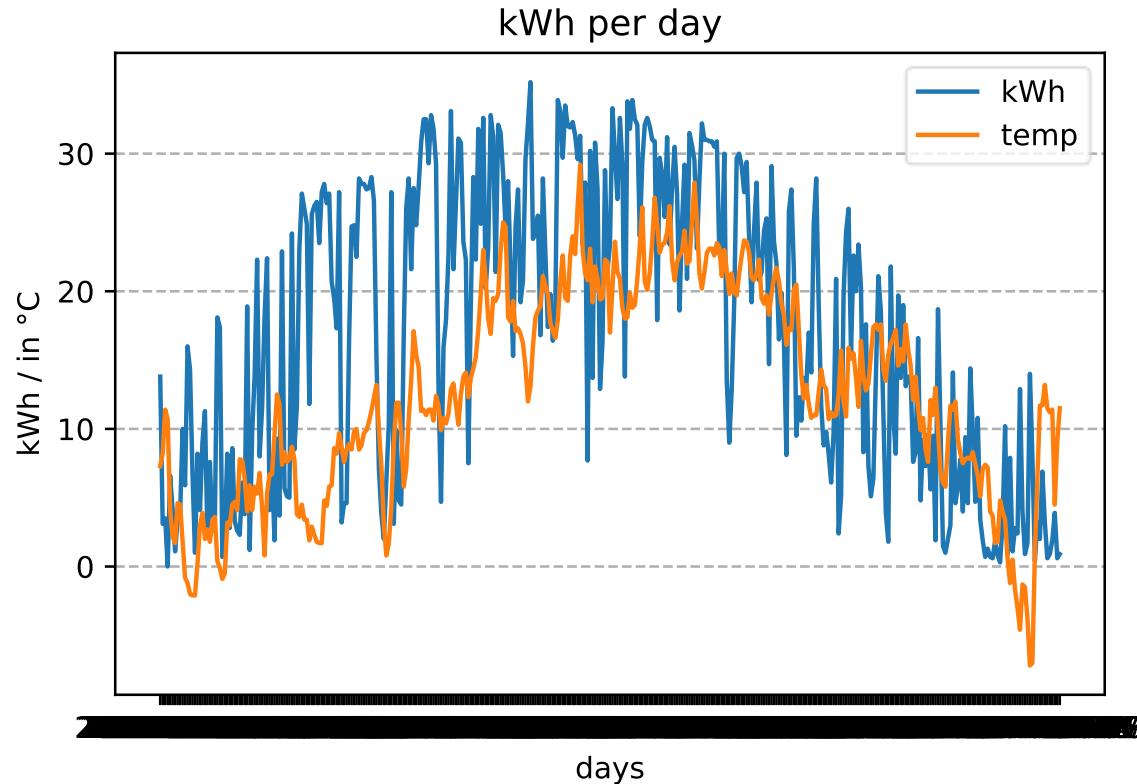
Legende:

- RSK = tgl. Niederschlagshöhe
- SDK = Sonnenscheindauer (Tag)
- SHK_TAG = Schneehöhe (Tag)
- NM = Tagesmittel des Bedeckungsgrades
- TMK = Tagesmittel der Temperatur
- TXK = Tagesmaximum der Lufttemperatur (2m Höhe)
- TNK = Tagesminimum der Lufttemperatur (2m Höhe)

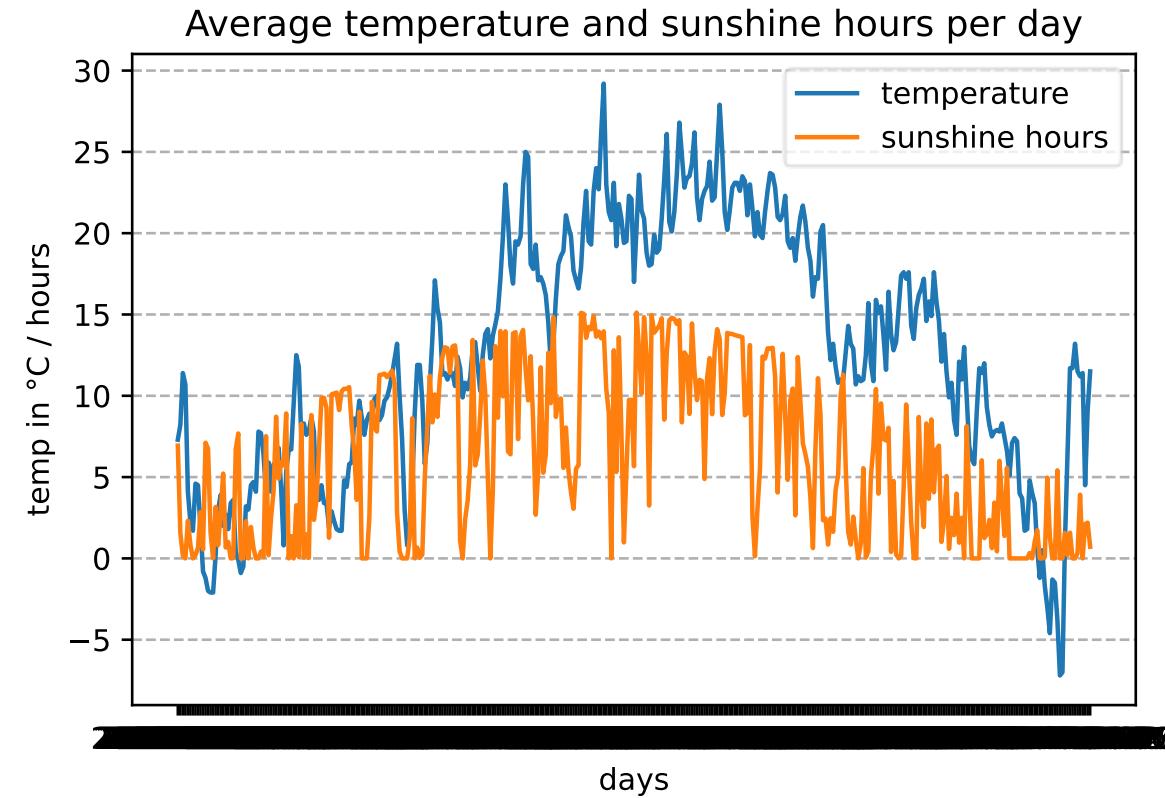
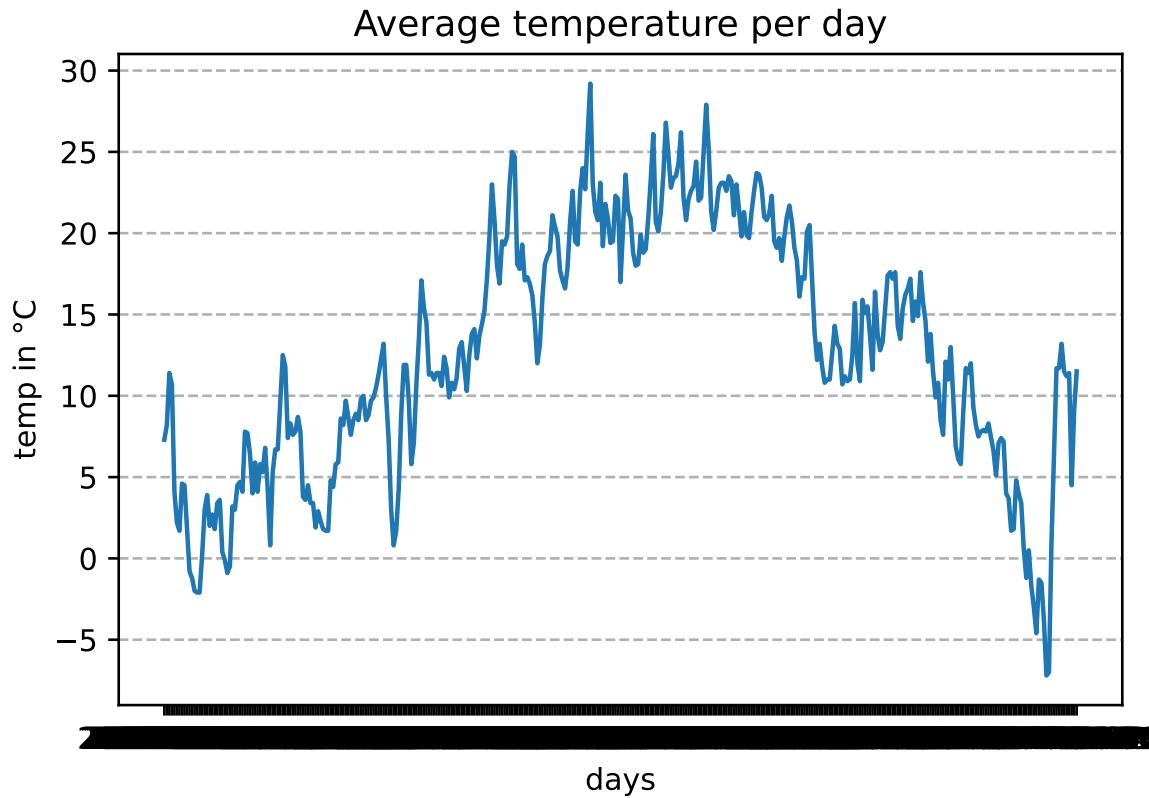
Daten – Diagramme kWh (1/2)



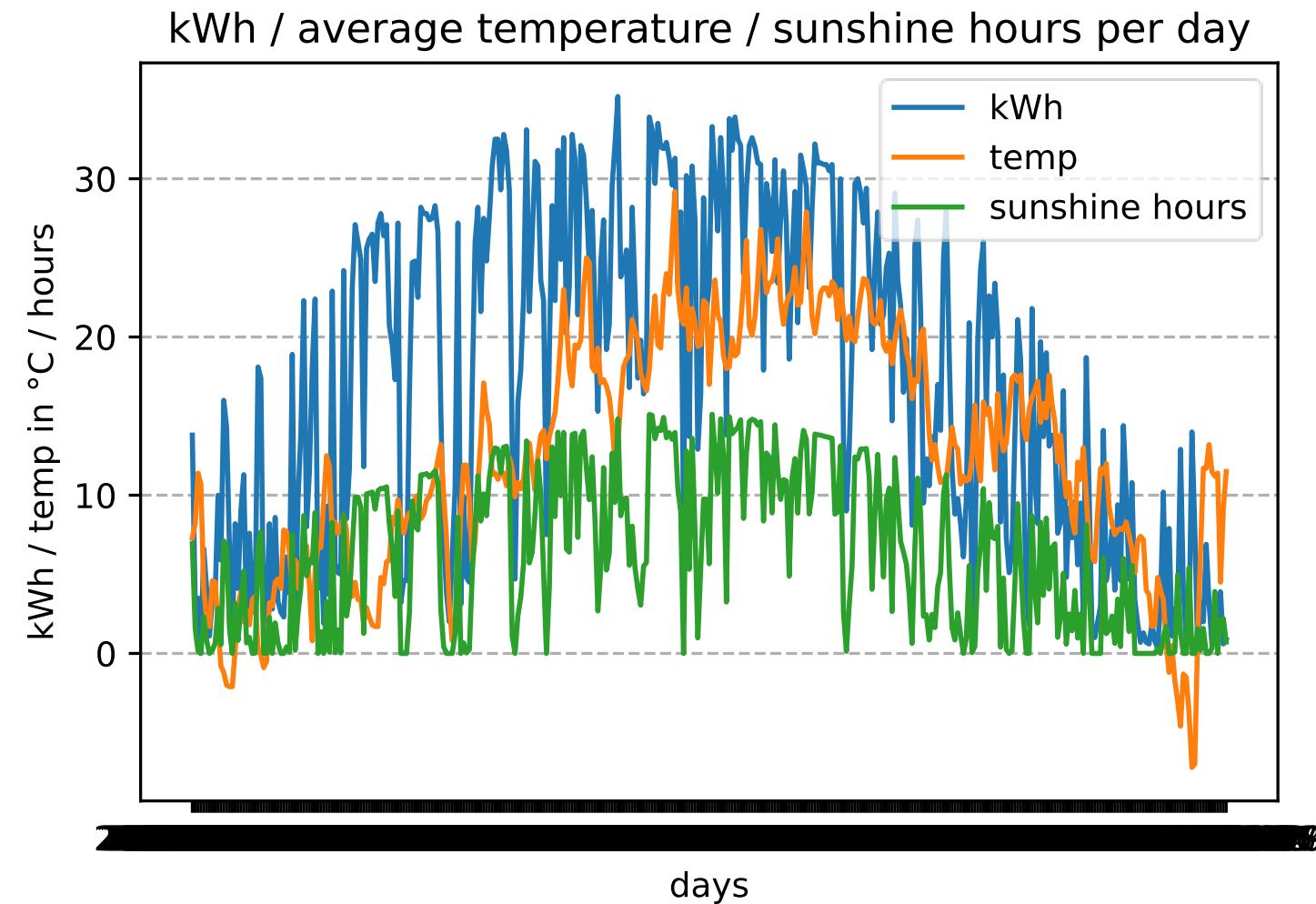
Daten – Diagramme kWh (2/2)



Daten – Diagramme Temperatur



Daten - Diagramm mehrerer Attribute



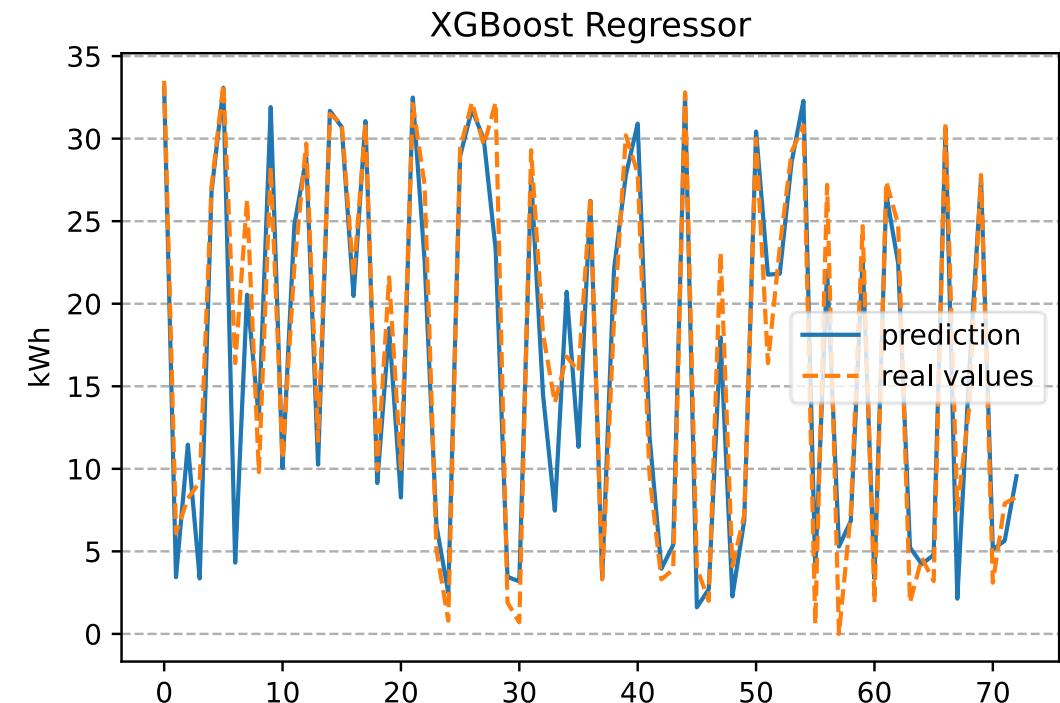
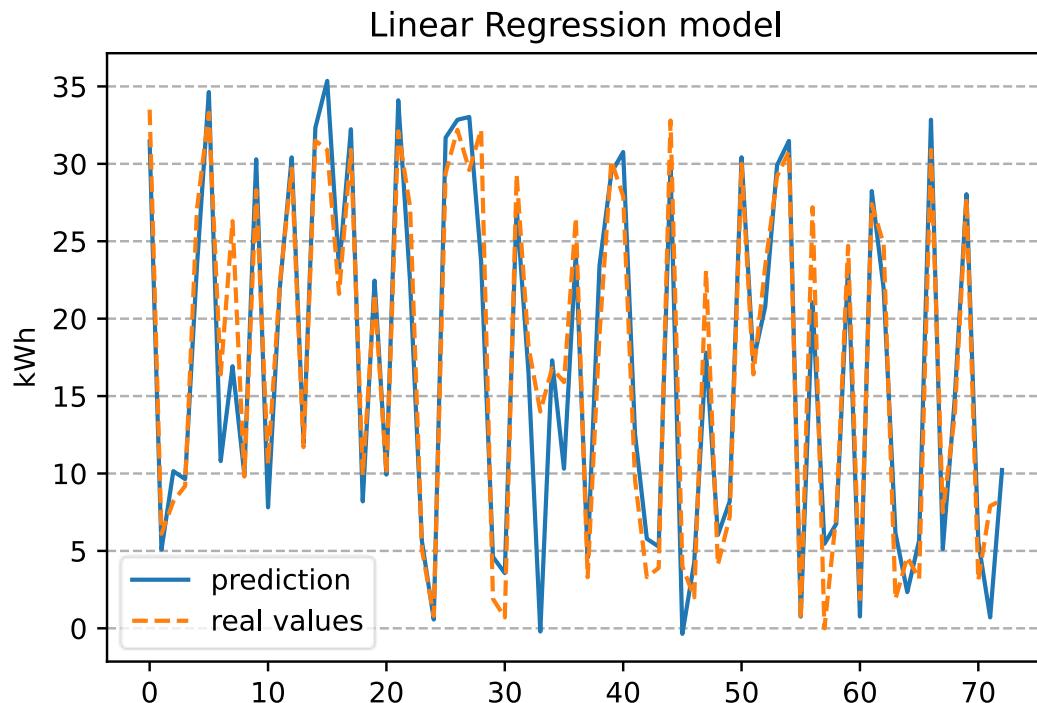
Regressions-Algorithmen

Algorithmen - Vergleich

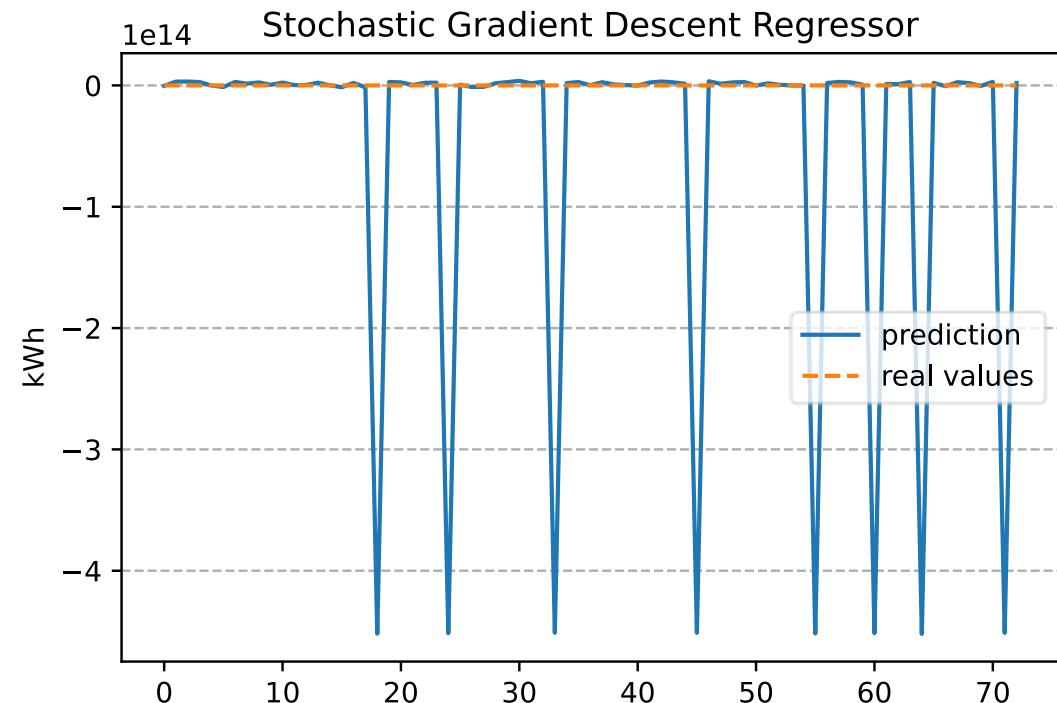
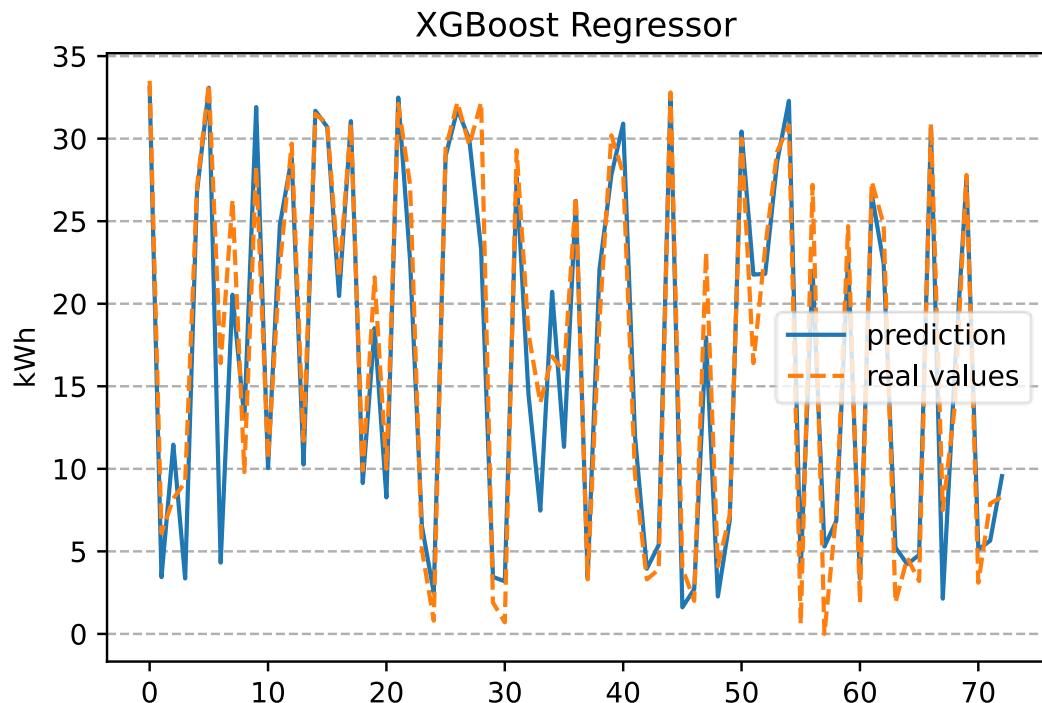
	Lineare Regression	XGBoost Regression	Stochastic Gradient Descent Regressor
MAE	2,48	2,27	73035006415375,55
MSE	11,86	9,97	2118917444741418975181 6683520,0
RMSE	3,44	3,16	145565017938425,69
Baseline (RMSE)	11,32	11,32	11,32

- Alle Algorithmen in der Basis Konfiguration
- Alle mit den gleichen Test- und Trainingsdaten modelliert (Daten der Version 2 mit der bereits erfolgten Attributsauswahl verwendet)
- Als Baseline wurde immer der Durchschnitt der produzierten kWh vorhergesagt
- Es wurde immer mit random_state / seed = 42 getestet, sodass die Ergebnisse vergleichbar sind.

Algorithmen - Diagramme



Algorithmen - Diagramme



XGBoost Regressor – Hyperparameter tuning

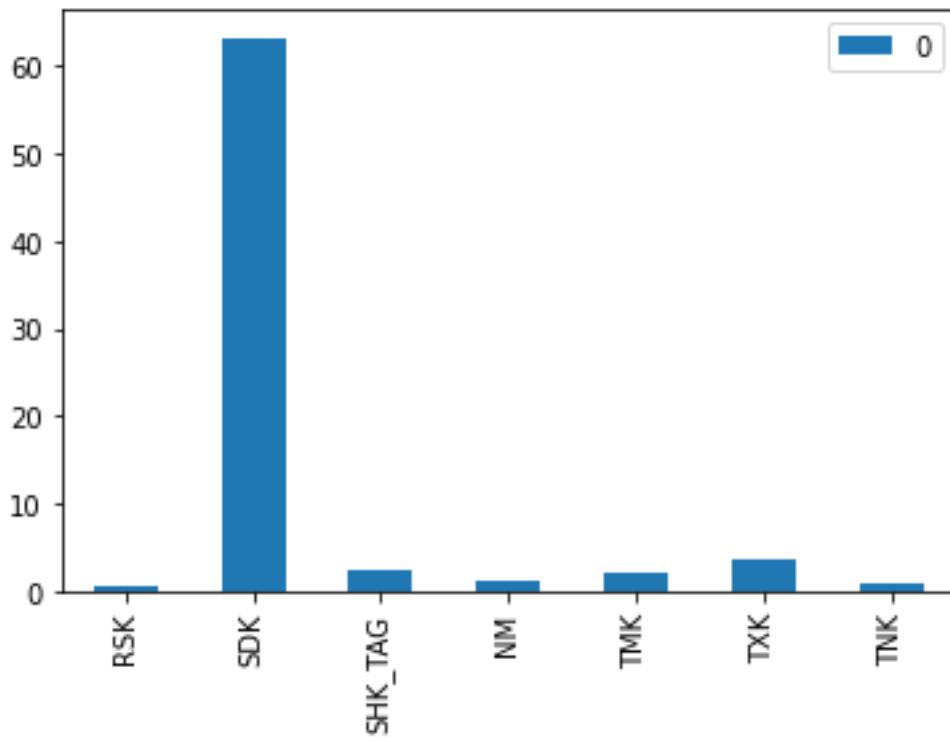
Folgende Parameter wurden verändert:

Parameter	Beschreibung	Wert	MAE	RMSE
-	-	-	2,27	3,161
eta	Learning rate	0.1	2,21	3,096
colsample_bytree	Anzahl an Attributen die je Baum verwendet werden	0.9	2,15	3,031
n_estimators	Anzahl an Bäumen in dem Ensemble	70	2,12	3,016
subsample	Anzahl an Daten-Reihen die je Baum verwendet werden	0.48	2,05	2,735

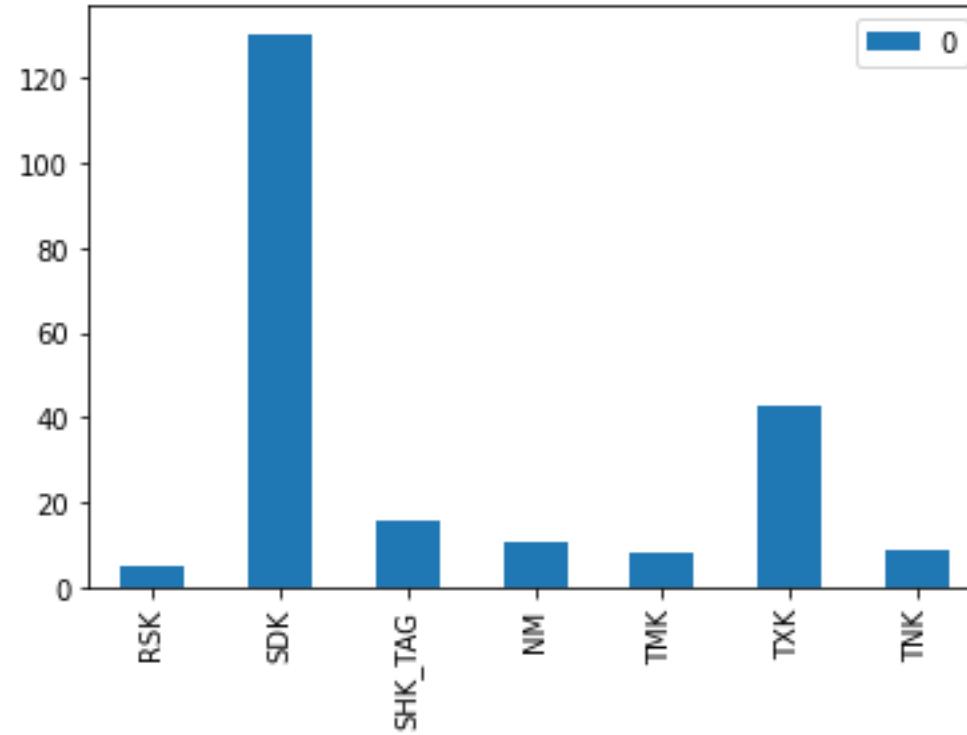
- Es wurde immer mit seed = 42 getestet, sodass die Ergebnisse vergleichbar sind.

XGBoost Regressor – Hyperparameter tuning

Feature importance vor dem
hyperparameter tuning:

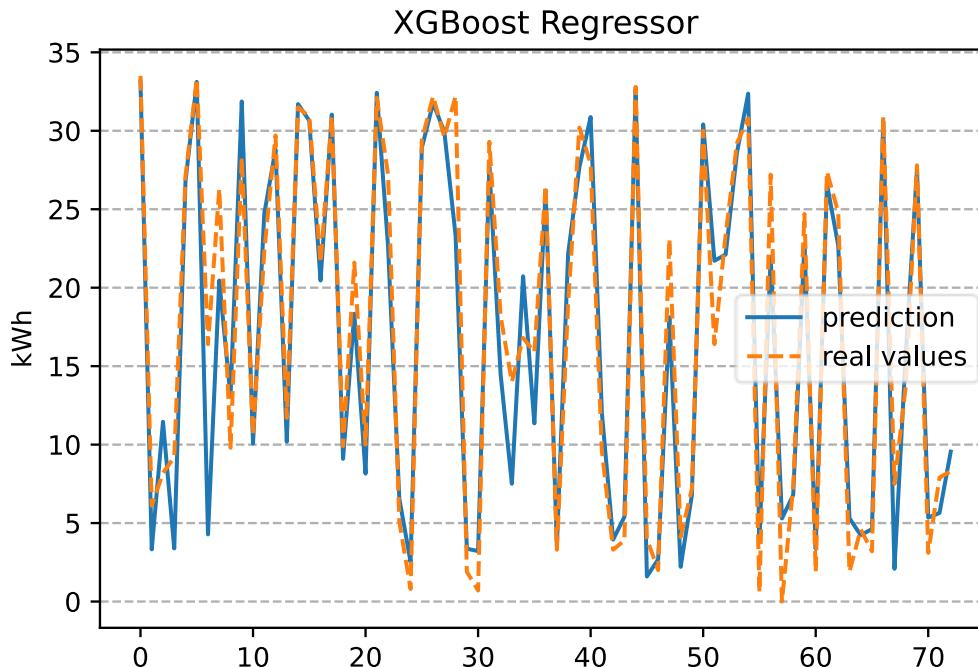


Feature importance nach dem
hyperparameter tuning:

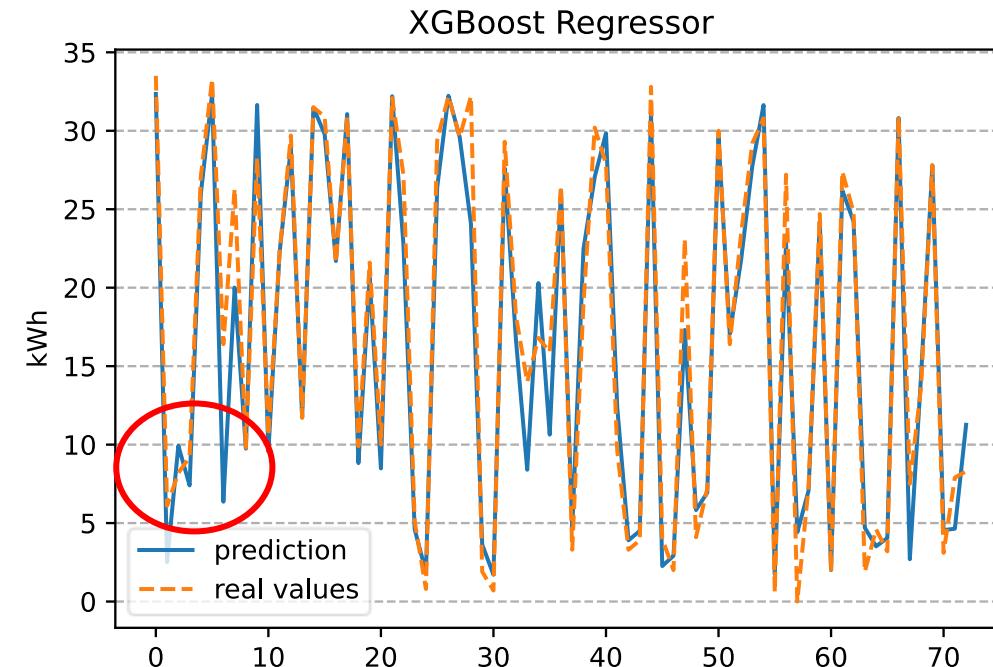


XGBoost Regressor – Hyperparameter tuning

Vergleich Vorhersage / reale Werte vor dem hyperparameter tuning:



Vergleich Vorhersage / reale Werte nach dem hyperparameter tuning:



Dashboard – Live Demo

Code

GitHub Link:

https://github.com/LeonRottler/Photovoltaic_plant_prediction



VORHERSAGE DER LEISTUNG EINER PHOTOVOLTAIKANALGE

Machine Learning

Leon Rottler – BAM – WS 22/23