technische universität dortmund

FiRRM Finance, Risk & Resource Management

Prognosis of the energy price mainly based on solar energy in Germany

A comparison of Random Forest, Gradient Boosting und XGBoost

Phillip Dürscheidt, Monika Pallasch, Laura Zammataro | Dortmund, 25.09.2024

Structure

- Introduction
- Data Collection & Cleansing incl. Visualization
- Three ways for regression - Comparison
- Regression Model and Results
- Analysis of our Results
- Finetuning
- Improvements

How was our procedure? - Introduction

1. Data collection & cleansing

2. Creation of three regression models

3. Analysis & Finetuning

Our Data → Focus: Solar energy
- Day-Ahead Total Load FC
- Actual Total Load
- Day-Ahead Solar generation FC
- Actual Solar Generation
- Sun hours
- Day-Ahead generation Last (w/o Solar)
- Electricity Prices

Our regression models:
- Random forest
- Gradient Boosting
- Extreme Gradient Boosting

→ Comparative indicator: R^2

Our analysis:
- Hyperparameter Tuning & Cross-Validation
- Feature Importance
- Adjustments of the Modell
  - Dummy variable
  - Adding explainable variable

Model assumptions

- Merit Order assumption
- Renewable energy → lowest cost, lower energy price[1]
- Solar power easy to relate to weather conditions
- Wind energy more complex

1 Vgl. Hirth (2022).

Data collection & cleansing

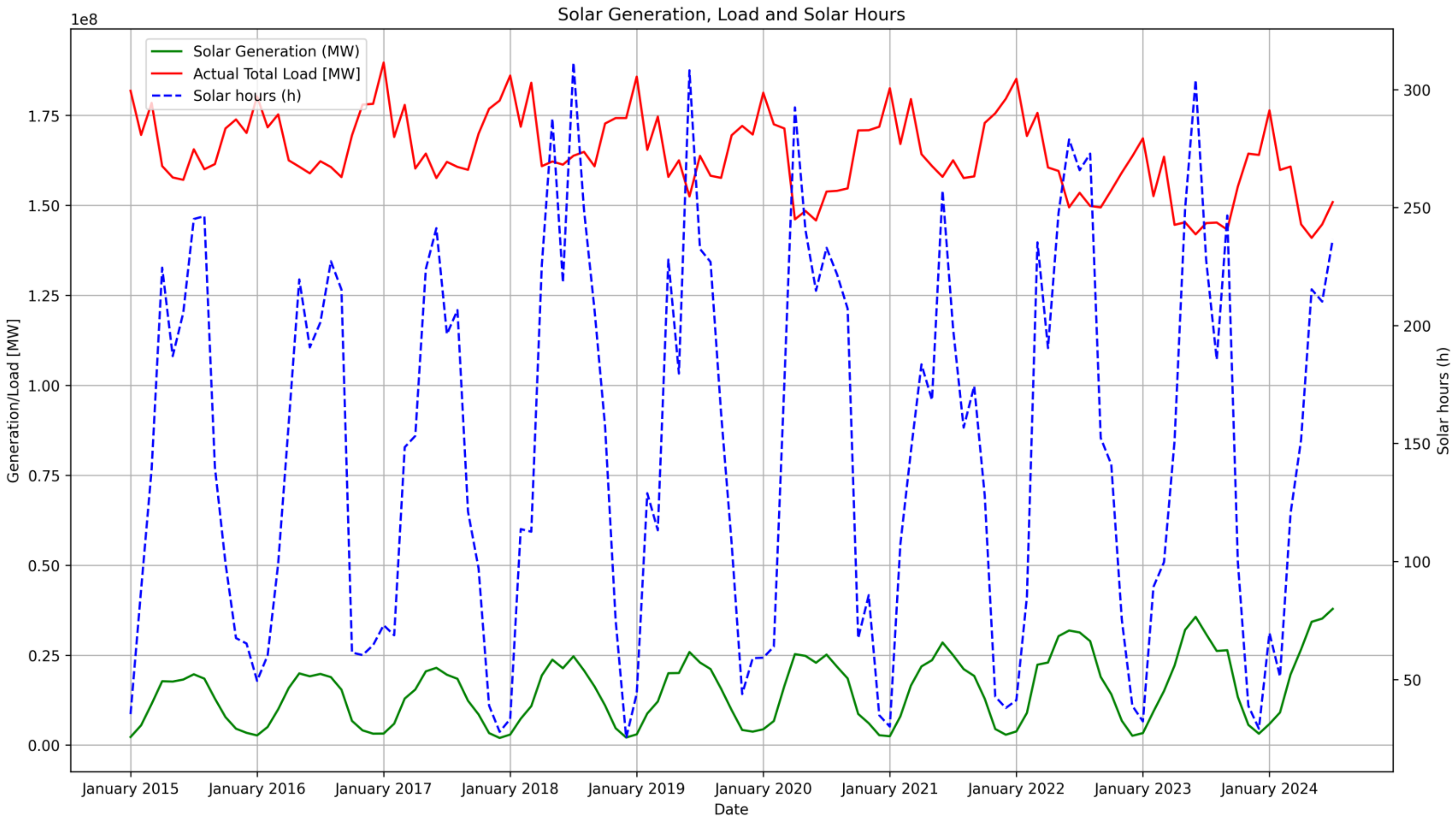
- Day-Ahead Total Load FC
- Actual Total Load
- Day-Ahead Solar generation FC
- Actual Solar Generation
- Sun hours
- Day-Ahead generation Last (w/o Solar)
- Electricity Prices

- Sources: Deutscher Wetterdienst, Frauenhofer ISE, entso-e Transparency platform, EMBER

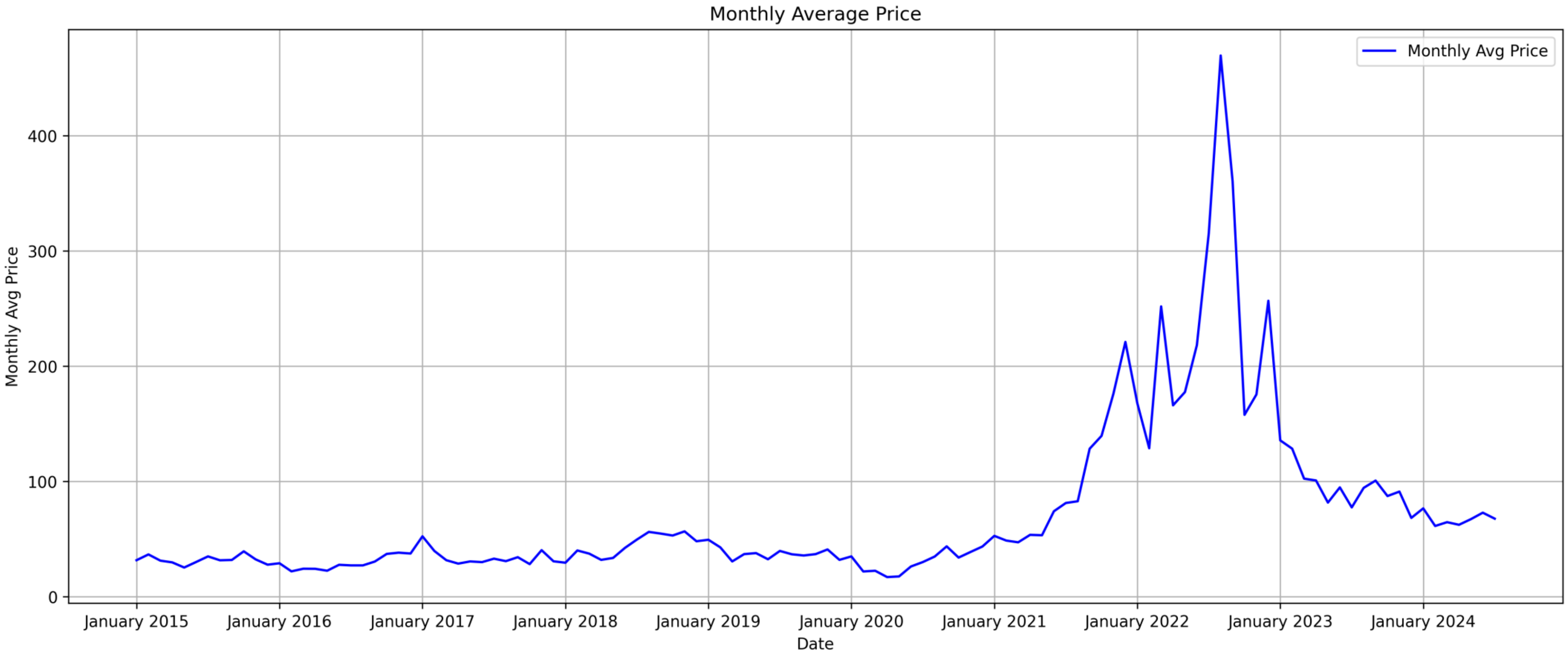
- Standardization → Monthly basis
- Reason: Sun hours just available per month or per hour
- Including all data in one file for further processing

- Data Visualization

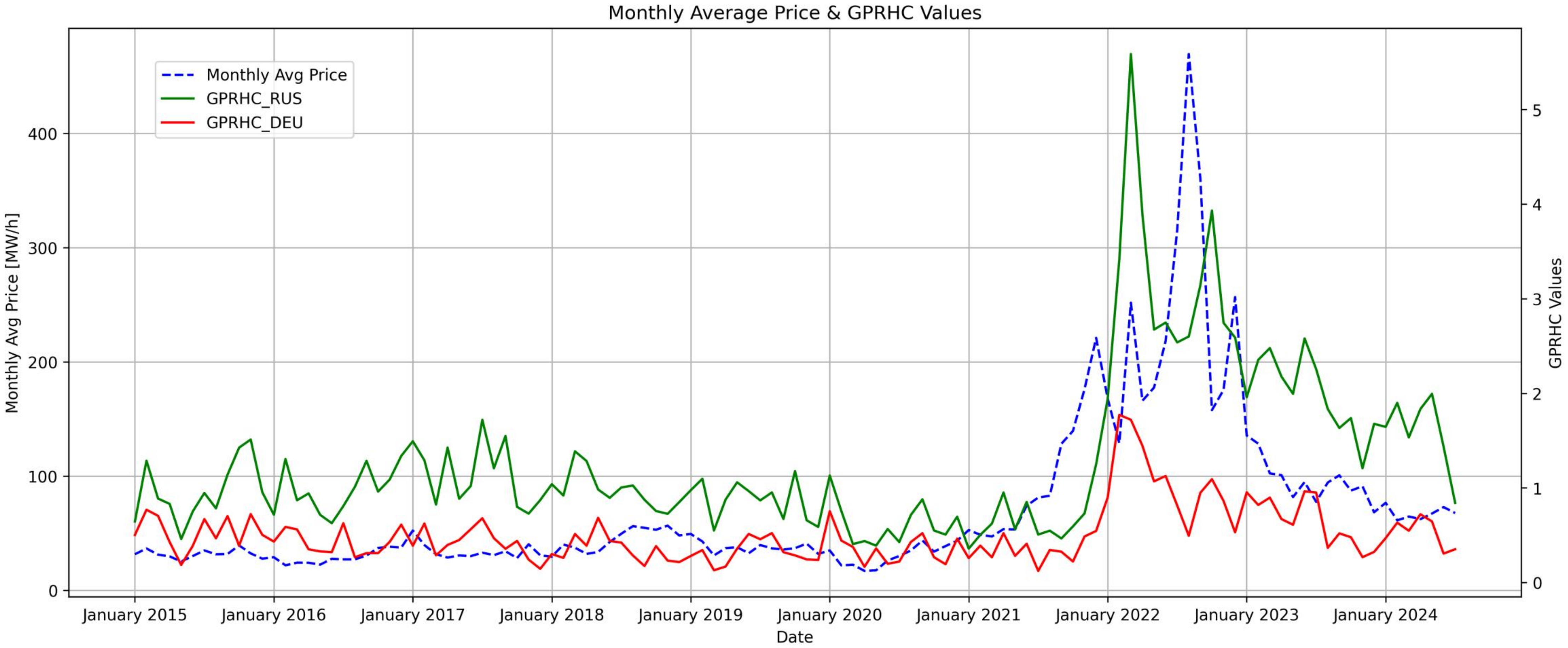
How did our data look like?



How did our data look like?



How did our data look like?



Comparison to other studies:

- **Sgarlato/Ziel (2023):** meterologic data improves electricity price forcasts, Germany, LASSO and matrices
- **Trebbien et al. (2022):** effect of renewable energy on the merit order, Germany, XAI machine learning, SHAP, GBT
- **Contreras et al. (2003):** autoregressive model, prediction based on historical data
- **Lucas et al. (2020):**explain and forecast the price behavior of a multi-variable regression model, Great Britain, GB, RF and XGBoost
- **Kumar Jana & Kumar Paul (2023):** hour ahead forecast of electricity prices, Spain, linear regression, XGBoost, CNN (Convolutional Neural Network), RNN (Recurrent Neural Networks), stacked RNN, LSTM, stacked LSTM etc.

Separation of the three used models:

Random Forest

- Combines multiple decsion trees to enhance accuracy and robustness
- Final predictions are made thorugh averaging
- Hyperparameter got tuned with gridsearch

Gradient Boosting

- Computes descision trees that learn from each other
- Computes a strong learner from multiple weak learners
- accuracy and performance are unmatched for tabular supervised learning tasks

XGBoost

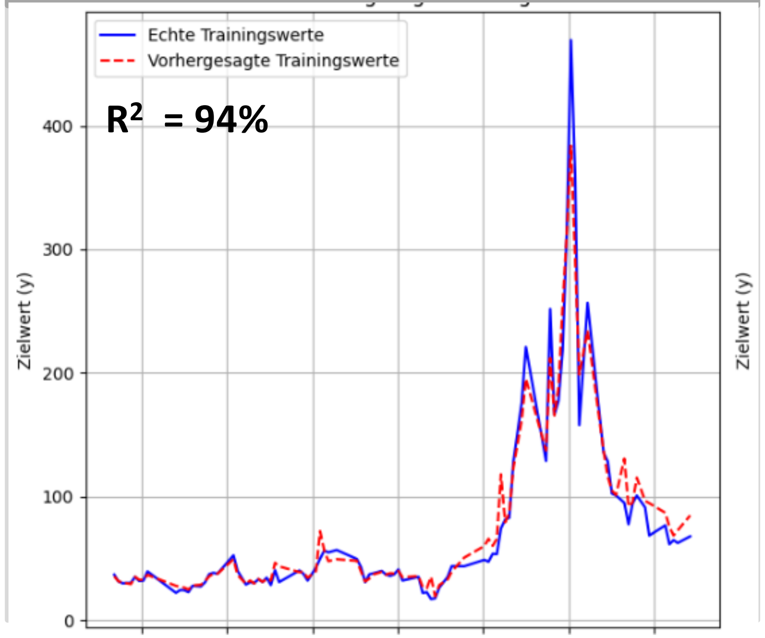
- Optimized version of gradient boosting
    - Speed
    - Efficiency
- Regularization of model complexity (L1 and L2 regularization)
- Parallelization (efficient computing time)
- Out-of-core computing (good for huge data sets)

What did our regression model look like?

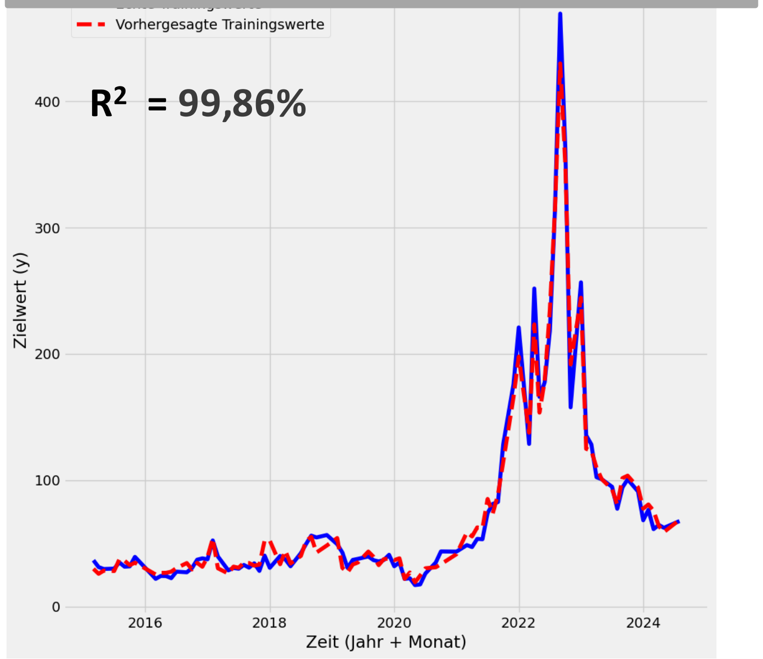
Electricity Price = b_0 + b_1 * Day-Ahead Load + b_2 * ACT Load + b_3 * Day-Ahead Solar gen. + b_4 * ACT Solar gen. + b_5 * Sun hours + b_7 * Day-Ahead gen. Last (w/o Solar) + b_8 * Date

Performance on training data:

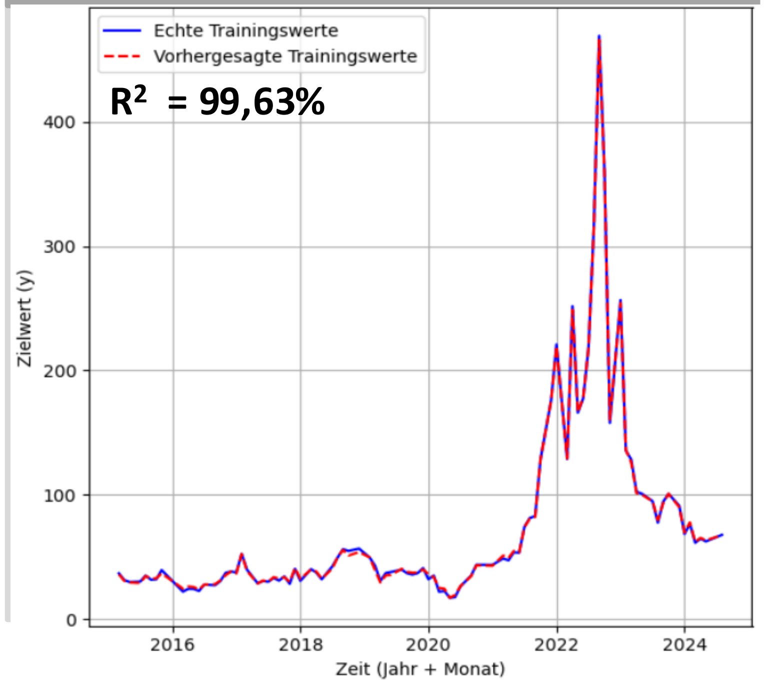
Random Forest



Gradient Boosting



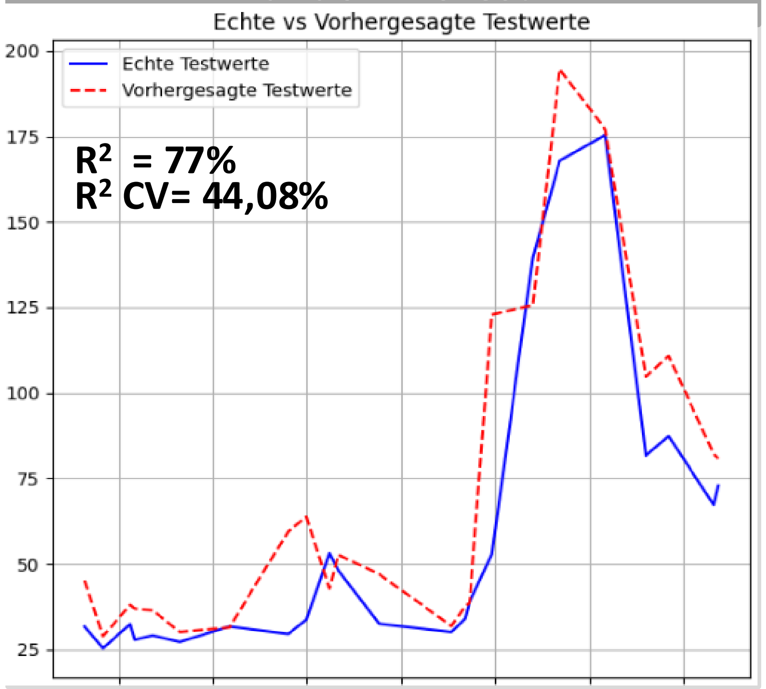
XGBoost



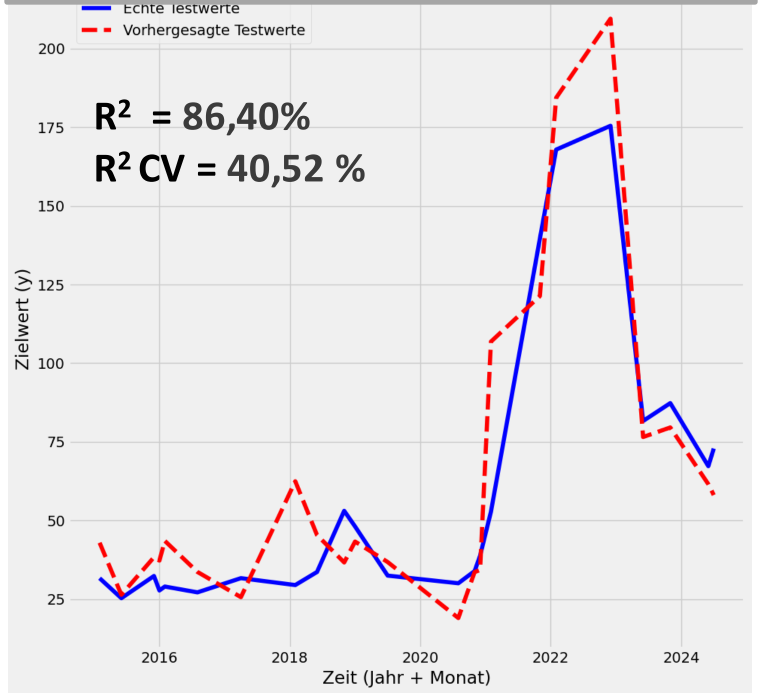
Our Results:

Performance on test data:

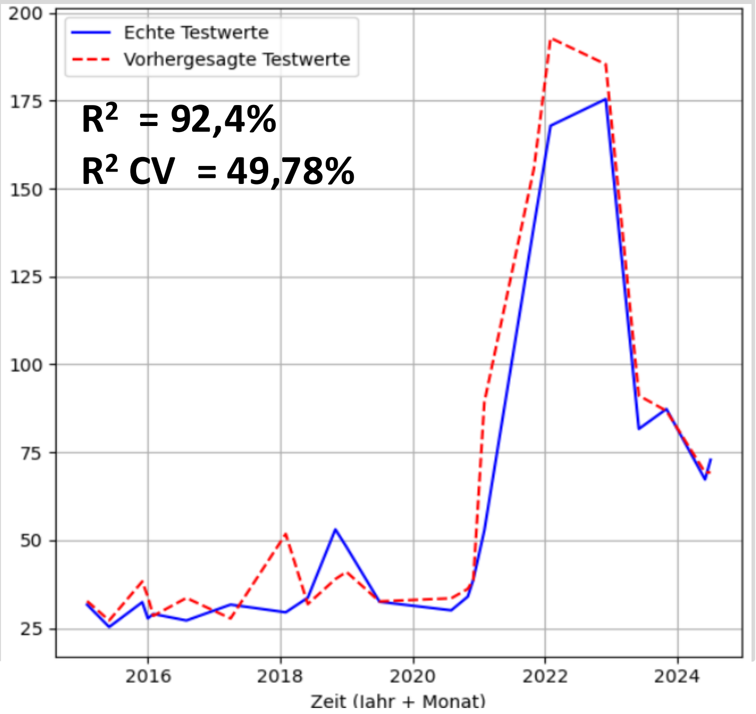
Random Forest



Gradient Boosting



XGBoost



XGBoost best Performance on test data due to highest R^2!

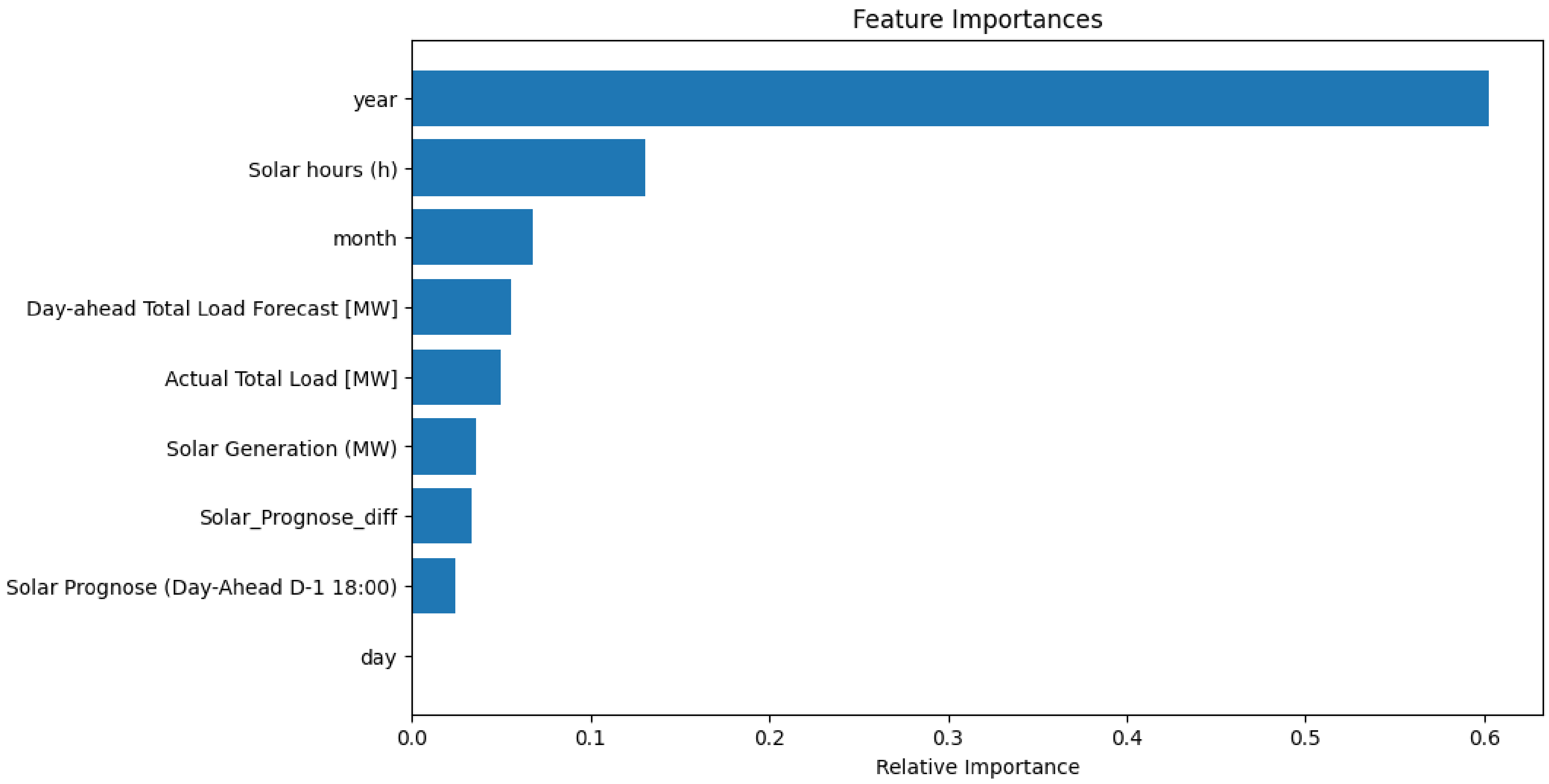
Further analysis: 1. Hyperparameter Tuning

	Random Forest	Gradient Boosting	XGBoost
Best Parameters found[1]:	Estimator: 20 Max depth: 15 Min samples split: 4 Max features: sqr	learning_rate: 0.1 max_depth: 3 n_estimators: 100	Learning rate: 0,2 Max_depth: 3 N estimators = 150
R^2 R^2 CV	R^2 = 53% R^2 CV = 53%	R^2 = 83,53% (from 86,40%) ↑ R^2 CV = 60,27% (from 40,52 %)	R^2 = 75,2% (from 92,4%) ↑ R^2 CV = 56,03% (from 49,78%)

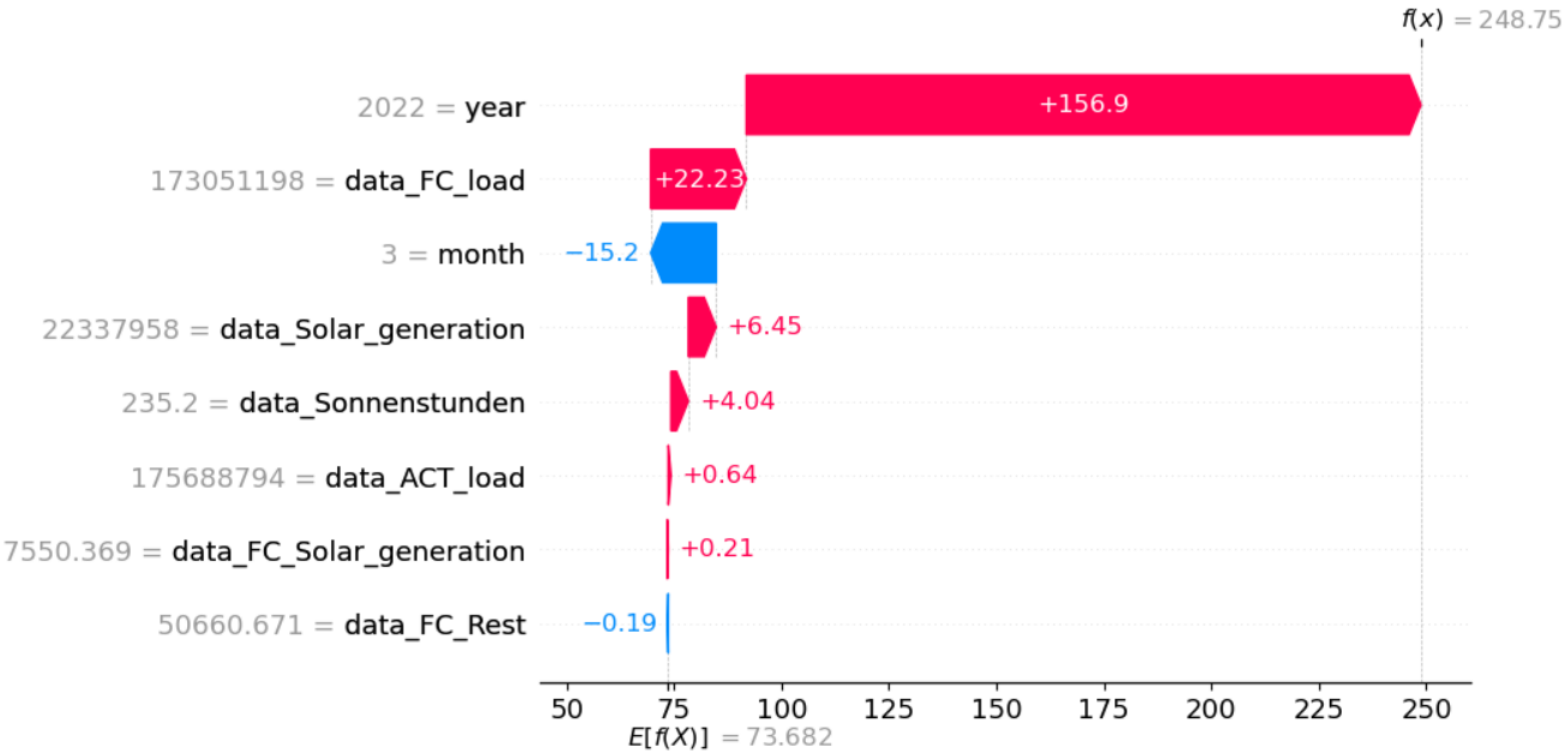
[1] used Gridsearch for optimized Parameters (Gökçe et al. (2022), P. 4)

Due to high dependence on test set R^2 CV optimized

Further analysis: 2. Feature Importance



Feature Importance - outlier



Comments:

- Analysis based on SHAP (SHappely Additive exPlanation) Values
  - Concept of game theory to identify the contribution of every single feature
- Year biggest impact on value
  - Mean prognosis: 73,668
  - Actual prognosted value of 248,75

  → Impact of 157 mainly caused due to year

Based on Lundberg (2017), p.1 f.

Adjustments

Initial Model:

Electricity Price = b_0 + b_1 * Day-Ahead Load + b_2 * ACT Load + b_3 * Day-Ahead Solar gen. + b_4 * ACT Solar gen. + b_5 * Sun hours + b_7 * Day-Ahead gen. Last (w/o Solar) + b_8 * Date

Dummy Variable

- signals the one-time effect of Ukraine war [1]
- Problem: not optimal for a small data basis
- XGBoost: Further CV-R^2 Improvement to 66,8% ↑
- Dummy for period: 2022-2023

Extracting effect by including GPR

- GPR (Geopolitical Risk Index)
    - War potencial by counting key words e.g. „War“ and assuming the connected riks

Too high feature importance of the date → biased by the ukrainian war (price detemermined by political regulations)

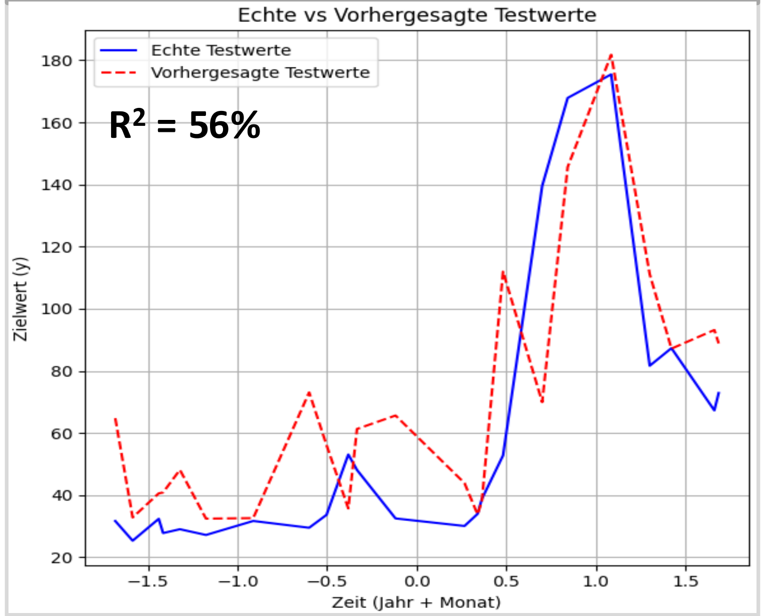
[1] Based on: Sgarlato/Ziel (2023), p.18

The adjusted regression model

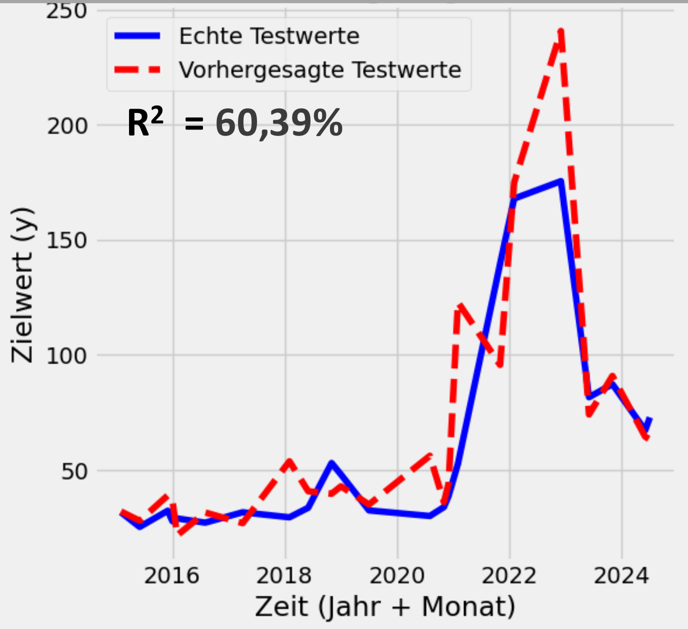
Electricity Price = b_0 + b_1 * Day-Ahead Load + b_2 * ACT Load + b_3 * Day-Ahead Solar gen. + b_4 * ACT Solar gen. + b_5 * Sun hours + b_7 * Day-Ahead gen. Last (w/o Solar) + b_8* Date + **b_9 * GPR Germany + b_{10}* GPR Russia**

Performance on test data:

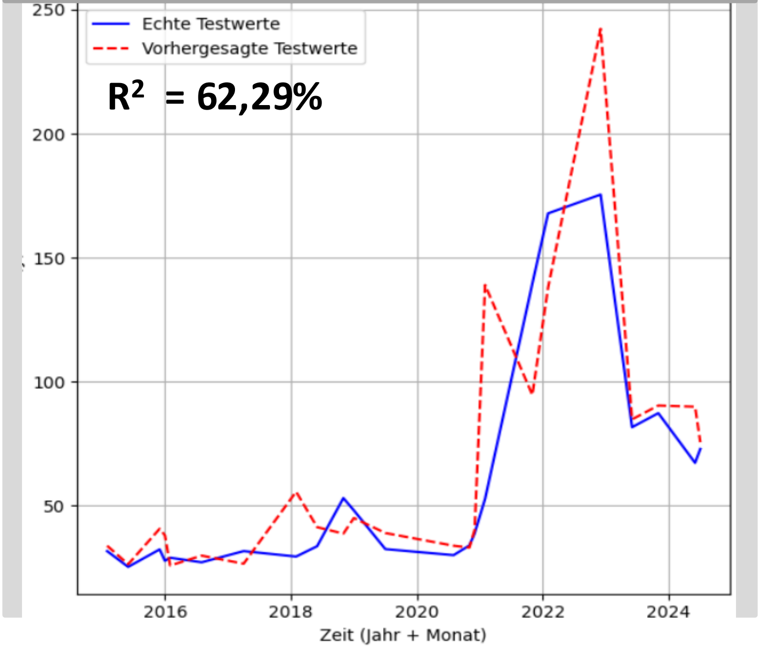
Random Forest



Gradient Boosting



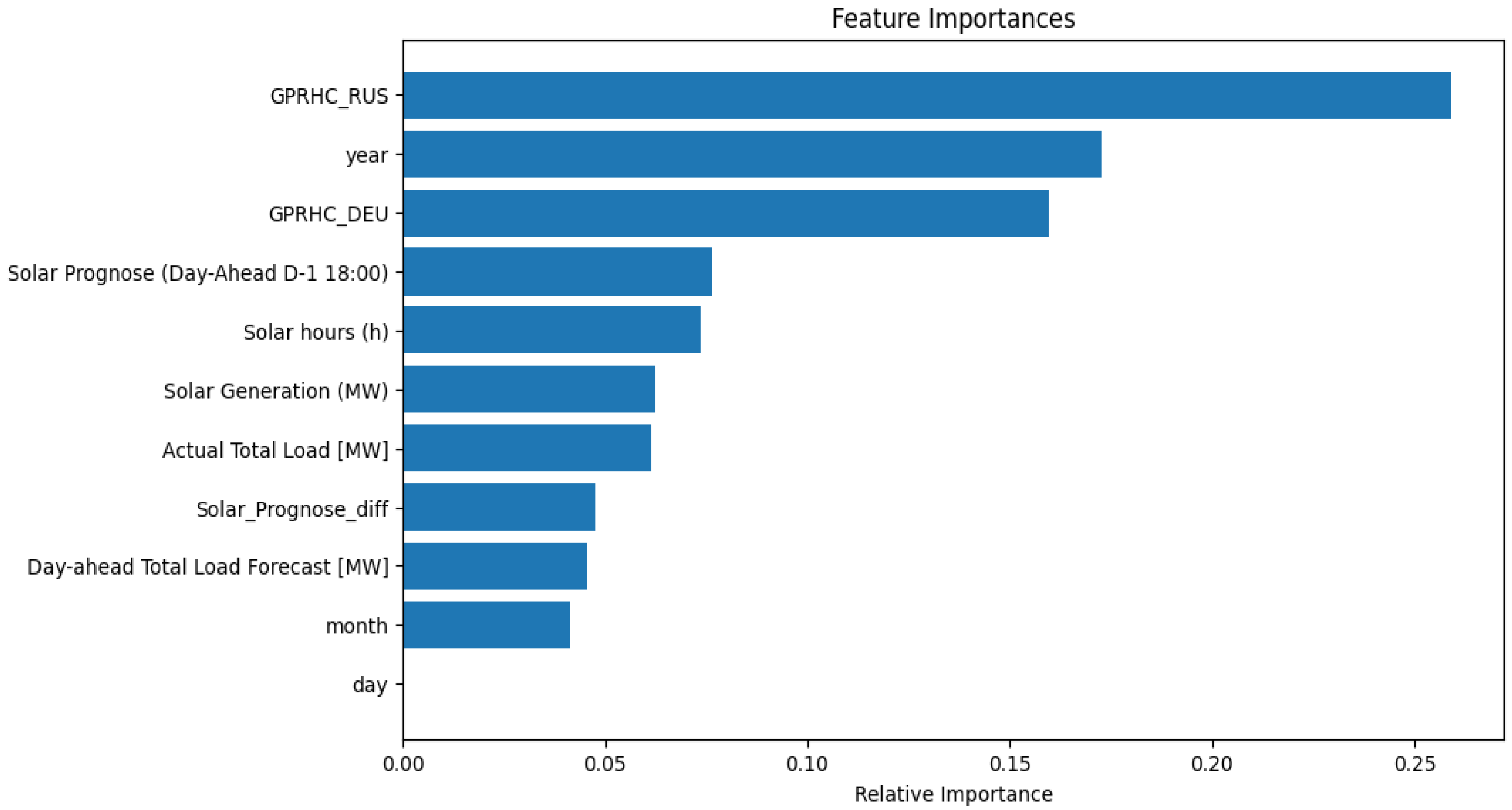
XGBoost



technische universität dortmund

FiRRM Finance, Risk & Resource Management

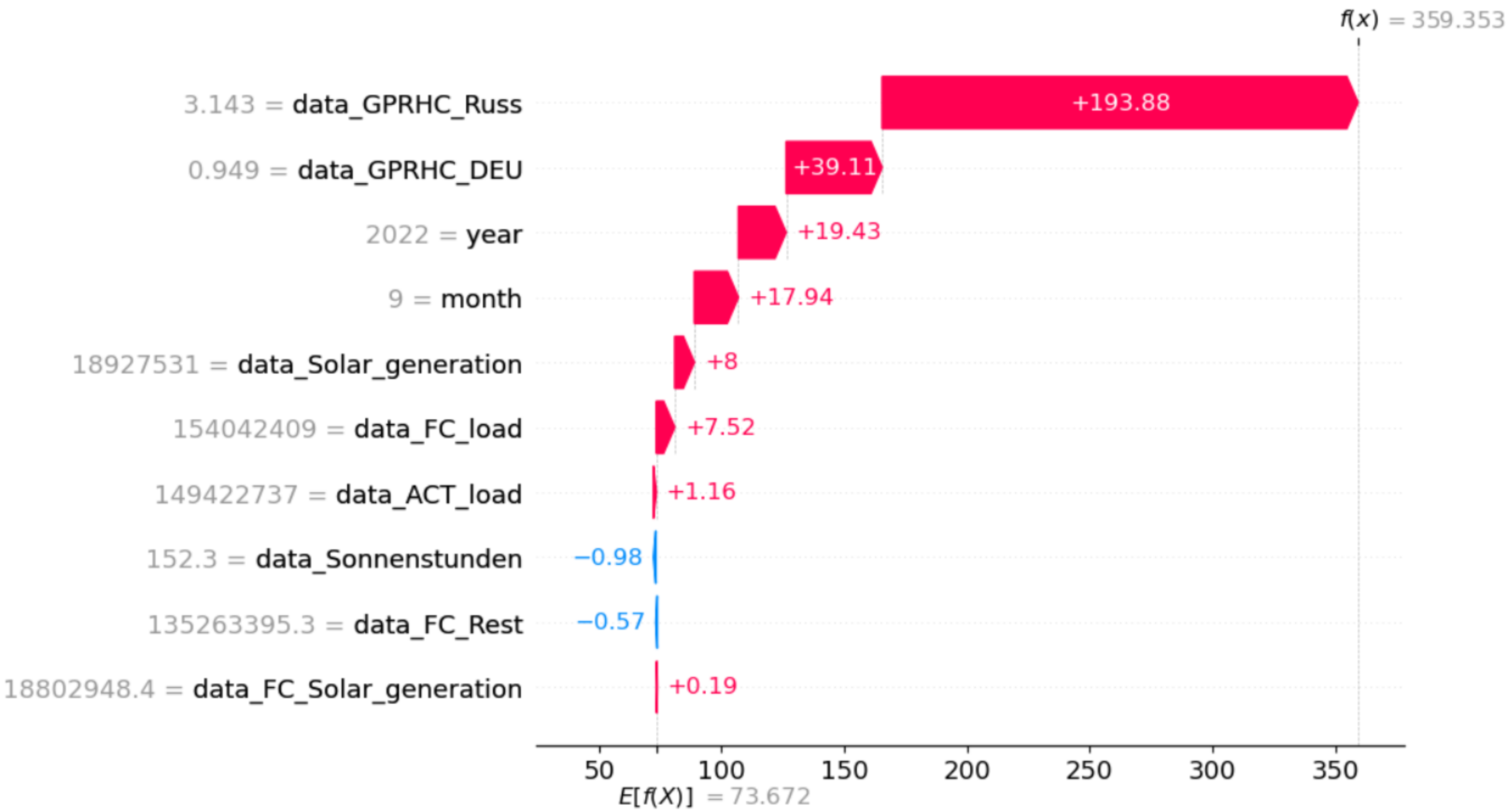
Analysis adjusted model: Feature Importance



Comments:

- The GPRHC_RUS variable shows the highest feature importance.
- The political effects were isolated from the 'year' variable.
- Despite this, the 'year' feature still holds high importance, likely due to other unaddressed factors.

Feature Importance – outlier adjusted model



Comments:

- Biggest impact not based on the date anymore
- Moved to GPR from Russia and Germany
    - Fits more real-world
    - Still not all impacts from date identified

Limitations / Improvements

Data:

- Daily / monthly basis not yearly
- More Data
- Include calendar and seasonal effects (holiday, weekend)
- Weather data like cloud cover, irradiance that influence PV plants, temperature influences load
- Location data, temperature higher in denser areas

Model:

- Finetuning for our adjusted Modell
  - Hyperparameter Tuning
  - Cross Validation
- Further extracting impacts out of the date
  - e.g. Inflation

General:

- Difficult to prognose electricity prices per month
- Market more determined by political regulations than weather or electricity demand, especially in the long term
- XGBoost regression achieved best results

Literature Sources

- Contreras, J., Espinola, R., Nogales, F., Conejo, A., (2003). ARIMA models to predict next-day electricity prices. In *IEEE Transactions on Power Systems 18 (*pp. 1014–1020).
- Gökçe, M. M., & Duman, E. (2022, December). Performance comparison of simple regression, random forest and XGBoost algorithms for forecasting electricity demand. In *2022 3rd International Informatics and Software Engineering Conference (IISEC)* (pp. 1-6). IEEE.
- Hirth, L. (2022). The Merit Order Model and Marginal Pricing in Electricity Markets. Neon energy. https://neon.energy/marginal-pricing (22.09.2024, 21:09).
- Kumar Jana, A., Kumar Paul, R. (2023). Performance Comparison of Advanced Machine Learning Techniques for Electricity Price Forecasting. In *2023 North American Power Symposium (NAPS).* IEEE.
- Lundberg, S. (2017). A unified approach to interpreting model predictions. arXiv preprint arXiv:1705.07874.
- Lucas, A., Pegios, K., Kotsakis, E., Clarke, D. (2020). Price Forecasting for the Balancing Energy Market Using Machine-Learning Regression. In *2020 Energies, 13, 5420.* doi:10.3390/en13205420
- Sgarlato, R., Ziel, F. (2023). The Role of Weather Predictions in Electricity Price Forecasting Beyond the Day-Ahead Horizon. In *IEEE TRANSACTIONS ON POWER SYSTEMS, Vol. 38, No. 3* (pp. 2500-2511).
- Trebbien, J., Rydin Gorjao, L., Praktiknjo, A., et al. (2022). Understanding electricity prices beyond the merit order principle using explainable AI. In *Energy and AI, Volume 13, July 2023, 100250.* Elsevier.