FORRS

Comparative Analysis of Machine Learning Models for Price Forecasting in the German Day-Ahead Market

Contents

1 Introduction

Electricity market participants rely heavily on accurate price forecasts to make informed operational and strategic decisions. Reliable forecasts support power producers in optimizing generation schedules, guide energy traders in formulating bidding strategies, and enable storage operators to plan charging and discharging cycles effectively.

Although the findings of this paper are transferable to other liberalized electricity markets, the focus lies on the German market. Electricity trading in Germany takes place across several market segments, including the control reserve market, the day-ahead market, and the intraday market. We focus in our analysis on the day-ahead market, as it represents the largest wholesale market segment and plays a central role in price formation. In 2024, the day-ahead auction accounted for 291.4 TWh of traded electricity, compared to 91.2 TWh on the continuous intraday market and 10.6 TWh in the intraday auctions (EPEX 2025).

In the day-ahead auction, participants submit quarter-hourly bids and offers for the following day. A market-clearing price is determined for each hour once bidding closes at 12:00, with results published shortly after. While quarter-hourly products were introduced to the day-ahead market in October 2025, this study focuses on 2024 data, when hourly products were the standard. This market design underscores the importance of timely and reliable day-ahead price forecasts, which are crucial for informed bidding and operational decision-making.

In recent years, the share of renewable energy in Germany's generation mix has increased substantially. While this transition supports decarbonization goals, it also introduces new challenges for electricity price forecasting. The weather-dependent nature of wind and solar generation, such as low output during calm or cloudy periods, leads to pronounced fluctuations in supply and, consequently, higher price volatility. Understanding and quantifying these effects are essential for improving forecast accuracy and market decision-making.

This paper compares three modeling approaches for forecasting day-ahead electricity prices: a traditional statistical model and two machine learning models, a tree-based approach and a neural network approach. Each model is evaluated under two training schemes: aggregated, where all 24 hourly prices are predicted jointly, and segregated, where individual models are trained for each hour. Their predictive performance is systematically compared to assess which approach offers the best balance between interpretability, economic value, and statistical accuracy.

The paper is structured as follows. Section 2 introduces the dataset and presents key descriptive insights. Section 3 outlines the modeling approaches and evaluation metrics, while Section 4 analyzes the empirical results. To illustrate the practical relevance of forecasting accuracy, in Section 5, we conduct a case study in which different price forecasting models are evaluated within an illustrative battery energy storage system optimization framework. This analysis demonstrates how forecasting performance can influence revenue potential. Section 6 concludes with a summary and outlook.

2 Data

This section provides an overview of the dataset, key descriptive insights that inform the forecasting models, and aims to understand the underlying behavior of the electricity price and its main drivers (demand, wind, and solar generation), as these dynamics directly shape the predictive performance and operational relevance of the models.

The data originates from ENTSO-E and covers two years of hourly day-ahead price observations as well as forecasted wind, solar, and electricity demand for Germany between January 2023 and December 2024. This two-year period captures a range of market conditions, including high renewable infeed phases and volatile price episodes, providing a representative basis for model development and evaluation. To ensure temporal consistency, the dataset was carefully adjusted for daylight saving time changes.

2.1 Endogenous Variable: DAA Electricity Price

Table 1 shows a summary statistic quantitatively describing some of the features of the applied data set. The average day-ahead auction price during the observation period was 86.83 €/MWh, with a median of 88.87 €/MWh. Prices typically fluctuated between 60 and 110 €/MWh, but extreme events occasionally pushed values as low as –500 €/MWh or as high as 936 €/MWh. Such spikes reflect the combined impact of renewable intermittency as well as market constraints and short-term imbalances.

Looking at the hourly average price pattern (Figure 1) shows a clear intraday cycle with two noticeable peaks, one in the morning around 8 a.m. and another one in the early evening around 8 p.m., driven by typical demand surges during these hours. Prices tend to be lowest during the night and midday, which can be attributed to reduced demand combined with higher renewable generation, particularly wind at night and solar during the day. This pattern highlights the need to account for intraday seasonality in forecasting models.

The series also shows a clear weekly seasonality (Figure 2), with lower prices on weekends and higher levels during weekdays. This reflects the recurring demand-supply cycle and highlights the importance of incorporating temporal patterns in model design. In practice, such weekly effects can significantly influence short-term trading strategies and storage optimization, underscoring the value of precise temporal modeling.

	mean	std	min	25%	50%	75%	max
Price	86.83	50.91	-500.00	62.92	88.87	112.34	936.28

Table 1: Descriptive Statistics of the Price Variable.

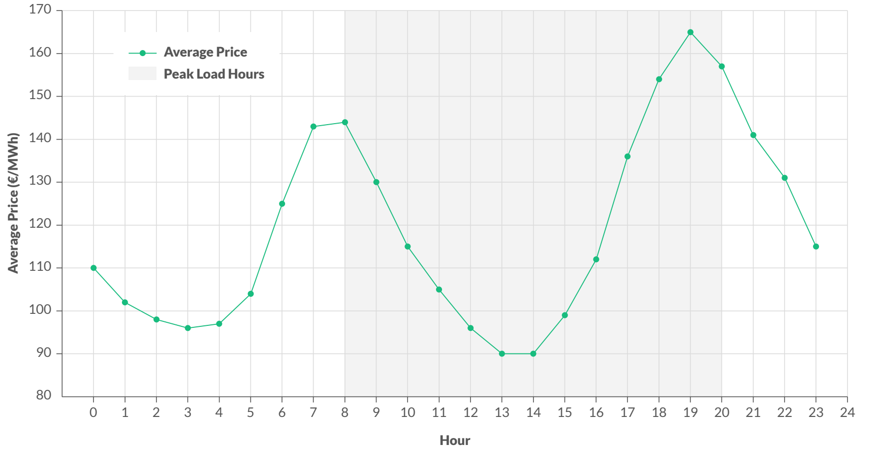

Figure 1: Average Day-Ahead Price for Each Hour.

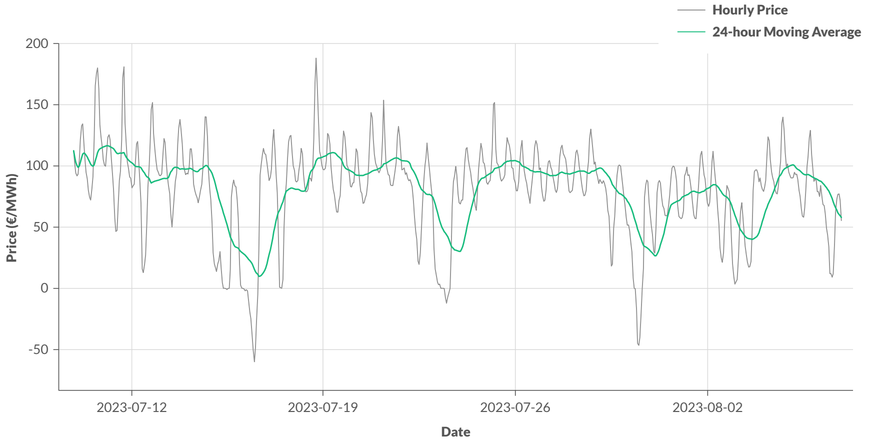

Figure 2: Day-Ahead Price and 24-hour Moving Average for the First Two Months.

2.2 Exogenous Variables: Wind, Solar, Load

For simplicity, we use only four different exogenous variables for our analysis. Table 2 summarizes the key characteristics of the exogenous variables used in the analysis. Among them, onshore wind generation exhibits the highest mean output (13.186 GW) and the largest variability, highlighting its strong influence on market conditions. Offshore wind contributes less on average (2.78 GW) but shows similar fluctuation patterns. Solar generation displays extreme variability, ranging from zero at night to peaks above 48 GW, with a very low median of 0.225 GW. In contrast, system load is comparatively stable, averaging around 53 GW within an interquartile range of roughly 45 GW to 61 GW.

	mean	std	min	25%	50%	75%	max
Wind Onshore	13.186	10.322	0.122	4.928	10.330	19.184	47.472
Wind Offshore	2.780	1.802	0.007	1.090	2.674	4.371	7.072
Solar	6.849	10.509	0	0	0.225	10.908	48.454
Load	53.099	9.126	30.893	45.604	52.836	60.628	74.042

Table 2: Descriptive Statistics for Wind, Solar, and Load Variables in GW.

The combination of highly variable renewable generation and relatively stable load indicates that these factors are key drivers of short-term price volatility. To quantify their relationship with electricity prices, an OLS regression was estimated, using normalized (0-1 scaled) inputs for better interpretability. The main results can be seen in Table 3.

	coef	std	t	P> |t|
const	84.22	0.65	129.18	0.000
Solar	-162.74	1.11	-146.08	0.000
Wind Onshore	-123.45	1.35	-91.45	0.000
Load	138.60	1.12	124.24	0.000
Wind Offshore	-29.69	1.13	-26.22	0.000
No. Observations: 17544			Adj. R-squared: 0.67	

Table 3: OLS Regression Results for the Relationship Between Electricity Prices and Normalized Exogenous Drivers.

All variables are statistically significant at the 1% level, confirming their strong link with electricity prices. Higher renewable generation, both solar and wind, tends to reduce prices, while higher load levels increase them. Offshore wind exerts a smaller but still significant price effect. With an adjusted R^2 of 0.67, the model explains a substantial share of the observed price variation, validating the inclusion of these exogenous variables in subsequent forecasting models.

3 Methodologies

Throughout this section, we focus on a real-valued time series denoted as $(Y_t)_{t\in\mathbb{N}}$. The primary objective is to provide predictions for each of the data points, represented as $\hat{Y}_t$. Each individual prediction, $\hat{Y}_t$, is derived from N inputs from the corresponding t-th data point, denoted $x_t = (x_{t,1}, \ldots, x_{t,N})$. These input features consist of exogenous variables, lagged values of those exogenous variables, and lagged values of the endogenous series Y_t.

To ensure realistic forecasting conditions, all models are evaluated on the out-of-sample period covering the year 2024. For each day within this evaluation horizon, the model is recalibrated using a rolling-window approach. This setup mimics a real-world forecasting scenario in which models are continuously updated as new data becomes available, thereby capturing the most recent market dynamics and improving short-term adaptability.

3.1 Data Preparation

Before model training, the input data is prepared to ensure consistency across the entire observation period. As the dataset does not contain any missing values, no additional handling is required in this regard. All timestamps are provided in Central European Time, including daylight saving time. Finally, where applicable, lags of each variable are included to capture temporal dependence.

3.2 Training, Validation, and Testing Procedure

Building on the prepared dataset, the forecasting framework employs a daily recalibration approach. For each forecasted day, the preceding 360 days are used for training, while the forecasted day serves as the test sample. For the machine learning methods, a separate validation set is created by splitting the 360-day training period into two subsets, where the first 75% of the data is used for training and the last 25% for the validation set. This enables the use of features such as early stopping. The hyperparameters of both machine learning models were tuned only once using data from the year 2023 and then kept fixed throughout the entire forecasting procedure. After each forecast, the entire forecasting window is shifted forward by one day, continuing until the end of the dataset on December 31, 2024.

3.3 Models

3.3.1 Statistical Model with Exogenous Inputs (Stat-EX)

This model extends classical time series analysis by combining autoregressive, moving average, and seasonal components with external regressors. It models how current values depend on past observations and forecast errors, filters out random noise, and captures recurring seasonal patterns such as weekly or yearly cycles. External drivers can be included as explanatory variables, making the model well-suited for forecasting tasks where both historical behavior and external influences matter. Despite its statistical complexity, the model remains interpretable and transparent for decision makers.

3.3.2 Decision Tree Based Model (Tree-ENS)

Decision trees are among the most intuitive machine learning models. They split data into successive decision rules, forming a tree-like structure that can be easily visualized and interpreted. Modern ensemble methods build upon this foundation by combining various such trees to improve predictive performance. One of the most effective approaches is gradient boosting (Friedman 2001), which sequentially combines many shallow trees, each one correcting the errors of the previous ensemble, to form a strong predictive model.

Figure 3 illustrates a simplified decision tree structure similar to those used in our forecasting framework. Each internal node represents a decision based on an input variable, such as solar generation, wind power, or recent price levels, while each leaf node corresponds to a predicted day-ahead electricity price. Although the actual model comprises hundreds of such trees, this schematic example highlights the interpretability and logical flow of tree-based methods

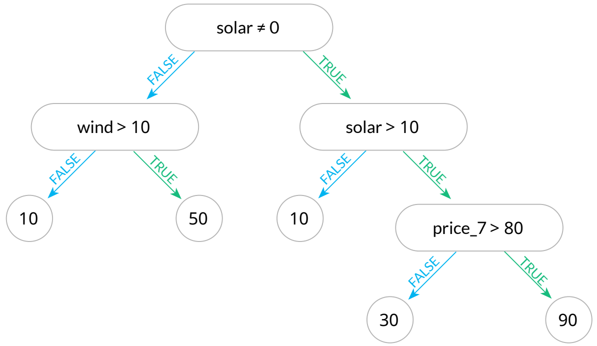

Figure 3: Schematic Illustration of a Decision Tree, Showing Exemplary Splits Based on Solar Generation, Wind Power, and Past Price Information.

3.3.3 Feed-Forward Neural Network (FFNN)

The network architecture illustrated in Figure 4 follows a feed-forward architecture consisting of an *input layer*, a *hidden layer*, and an *output layer*. The input layer receives a vector of features $x_t = (x_{t,1}, x_{t,2}, \ldots, x_{t,N})$, which are first standardized using a MinMax scaler before being compared with a set of representative centers in the hidden layer. Each hidden unit center c_j determines the region of input space to which it is most sensitive, enabling the model to represent complex and nonlinear relationships between the inputs and the target variable. Throughout the process of determining the centers, the number of hidden neurons is determined automatically. The output layer then aggregates these values through a weighted combination to produce the final prediction $\hat{Y}_t$. The forward-only flow of information allows the model to be trained efficiently while retaining flexibility to approximate highly nonlinear mappings. Recurrent neural networks specifically designed for sequential data, such as LSTMs, are not the subject of this paper, as the FFNN architecture demonstrated worse performance in our forecasting experiments and is less computationally demanding than the aforementioned.

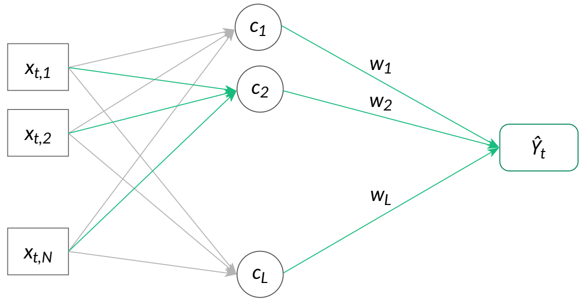

Figure 4: Structure of a Feed-Forward Neural Network. Each Hidden Node Applies a Localized Activation Function Centered at c_j to Transform the Input.

3.4 Model Characteristics

Table 4 shows a summary of important characteristics. Both the sequential learning process of gradient boosting in Subsubsection 3.3.2 and the localized responses in Subsubsection 3.3.3 are particularly valuable when system behavior is highly non-linear and cannot be adequately captured by simpler statistical approaches. Compared to the Tree-ENS, the feed-forward neural network model offers greater flexibility in capturing complex interactions seen in the price data in Subsection 2.1. However, it is also computationally demanding and the least interpretable.

	Stat-EX	Tree-ENS	FFNN
Non-linear	–	+	++
Interpretable	++	–	––
Comp. Time	–	+	0

Table 4: Summary of Model Characteristics.

3.5 Segregated vs Aggregated

To assess the effect of data granularity on model performance, a segregated and an aggregated setup are implemented as two alternative training approaches.

In the segregated approach, each product is modeled separately to account for its distinct characteristics. Hyperparameters are determined individually for each product, and the remaining model parameters are trained exclusively on the corresponding subset of data. This results in specialized models tailored to each product's specific dynamics, but may increase the risk of overfitting due to the limited number of observations per model.

In contrast, the aggregated approach combines all product data into a single, unified training dataset. Here, hyperparameters are optimized once for the complete dataset, and the remaining parameters are trained jointly on all available observations, effectively increasing the data volume by a factor of 24 compared to the segregated case. This produces smoother and more generalized parameter estimates that are identical across products, improving robustness but potentially reducing the model's ability to capture product-specific patterns.

The comparison between these two strategies highlights the trade-off between model specialization and generalization. Segregated models can better adapt to local variations, whereas aggregated models benefit from broader data coverage and reduced variance.

3.6 Performance Measures

The performance of all models is evaluated on out-of-sample data from the year 2024. Three complementary performance measures are employed: the Root Mean Squared Error (RMSE), the Mean Absolute Error (MAE), and the Forecasting Efficiency Ratio (FER). All models are trained using the Mean Squared Error (MSE) loss, which directly corresponds to the RMSE metric, while MAE and FER are used only as additional evaluation measures.

The RMSE and MAE are standard statistical metrics that quantify the accuracy of most time series forecasts, and are valid measurements in the case of German electricity prices. RMSE penalizes larger forecast deviations more strongly, making it sensitive to outliers, while MAE provides an interpretable measure of the average absolute forecast error. Together, they give a balanced view of each model's predictive accuracy.

In addition to standard statistical evaluation, we assess the quality of our forecasts through a BESS case study. Specifically, we measure how efficiently the forecasts translate into revenue when used by a BESS asset in day-ahead electricity market bidding. This is done by comparing the actual revenues achieved using our forecasts to the maximum possible revenues that could be obtained with perfect foresight, hence calculating FER.

4 Econometric Analysis

This section compares the forecasting performance of the models introduced in Section 3, evaluating both aggregated and segregated approaches in terms of overall and product-specific accuracy.

	Segregated		Aggregated	
	RMSE	MAE	RMSE	MAE
Tree-ENS	35.31	23.50	27.94	18.08
FFNN	31.39	18.30	27.54	17.05
Stat-EX	31.33	16.12	27.16	16.02

Table 5: Average Test RMSE and MAE of Segregated and Aggregated Models.

Table 5 shows that aggregated models consistently outperform segregated models across all model types. Stat-EX achieves the lowest errors overall, followed by FFNN, while the Tree-ENS performs the weakest. Similar patterns are observed on the training and validation sets. The improved performance of aggregated models is likely due to their training on larger and more diverse datasets, which reduces the potential of overfitting and improves generalization.

Figure 5 illustrates the MAE for each product. Forecast errors are generally lowest during nighttime hours (22:00-06:00) and highest during late afternoon (16:00-20:00), which stems from the high volatility of solar energy production during daytime. Aggregated and segregated Stat-EX models consistently achieve the best hourly performance, while FFNN performs competitively at night. Both Tree-ENS models show the weakest performance, with the segregated variant performing worst.

These results suggest that leveraging the strengths of different models for different products could improve overall accuracy. For instance, using aggregated Stat-EX forecasts for early-morning hours and segregated FFNN forecasts for late-night hours may further enhance predictive performance.

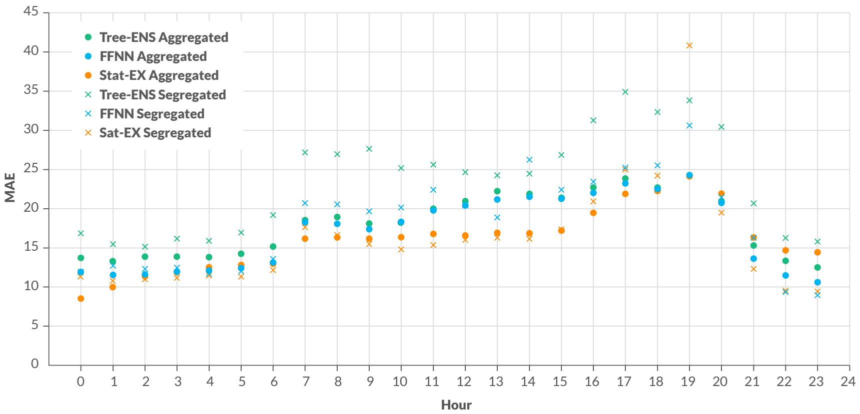

Figure 5: Hourly MAE Comparison of Aggregated vs. Segregated Models.

5 Use Case

To evaluate the practical relevance of the forecasting models, we apply the publicly available *FlexIndex* framework (*FlexIndex 2025*) as an application-oriented benchmark of forecast performance. Unlike conventional error metrics such as RMSE or MAE, this approach assesses the economic value of a forecast by simulating its performance in actual trading decisions. Specifically, it estimates the profit that a BESS could obtain in the day-ahead market when bidding based on forecasted price trajectories and compares it to the hypothetical revenues that would be achieved under perfect foresight.

It maximizes daily trading revenues subject to standard operational constraints on energy capacity, power limits, and state-of-charge dynamics. The optimized schedule is then evaluated ex post using the realized market prices to obtain the actual profit, thereby linking forecast accuracy directly to economic outcomes.

The efficiency ratio quantifies the economic impact of the forecast quality as the ratio between the revenues achieved and with forecast-based trading and those attainable under perfect foresight:

$$\mathrm{FER} = \frac{\Pi_{\mathrm{forecast}}}{\Pi_{\mathrm{perfect}}} * 100\%.$$

A value of FER=100% represents perfect predictive performance, while a lower values indicate efficiency losses caused by forecast errors.

The optimization is repeated for each day of 2024 using the identical model and battery specification. Aggregating the resulting daily profits yields annual efficiency ratios that quantify how effectively each forecasting model translates predictive information into market value, providing a practically interpretable measure of forecast performance. As shown in Table 6, the Stat-EX Aggregated model achieves the highest annual revenues among all evaluated forecasting methods, corresponding to an efficiency ratio of 95% relative to the perfect-foresight benchmark. This indicates that the model captures market dynamics with high accuracy, enabling trading decisions that closely approximate the theoretical optimum.

In real-life BESS trading, computational efficiency can be key, usually depending on the design of the market. However, for the day-ahead market as an auction market, speed can be considered less relevant. Since price prediction is a recurring process, it is necessary to regularly update the hyperparameters using systematic tuning methods such as grid search, random search, or Bayesian optimization.

Forecasting Model	FER [%]
Perfect foresight benchmark	100
Stat-EX Agg	95
FFNN Agg	92
Stat-EX Seg	90
Tree-ENS Agg	86
Tree-ENS Seg	86
FFNN Seg	84

Table 6: Comparison of Model Performance Based on Efficiency Ratios in 2024.

6 Conclusion

The objective of this study was to forecast DAA market prices to support BESS operations and decision-making. Several model families were evaluated, including classical statistical, neural-network, and tree-based models, each tested in both aggregated and segregated setups.

From an econometric perspective, the results were consistent and conclusive. The aggregated models clearly outperformed segregated models across all configurations. This advantage likely stems from the fact that aggregated approaches are trained on larger and more diverse datasets, which reduces overfitting and enhances generalization. Among the individual models, Stat-EX achieved the lowest overall errors, followed by the FFNN, while Tree-ENS showed comparatively weaker performance.

Furthermore, the case study demonstrates that forecasting accuracy has a direct and measurable impact on the economic value of trading decisions. Among all evaluated models, the aggregated Stat-EX model achieved the highest performance in terms of the efficiency ratio derived from the battery trading simulation. Aggregated model variants consistently outperformed their non-aggregated counterparts, confirming that the combination of multiple information sources or model components leads to more robust and economically valuable predictions. The second-best performing model in terms of efficiency ratio, aggregated FFNN, achieved comparably strong results, further supporting the conclusion that ensemble and aggregation strategies demonstrate both predictive accuracy and operational profitability. In summary, the results highlight the importance of evaluating forecast models not only through statistical error metrics but also through their realized market performance when applied in realistic trading scenarios.

Looking ahead, several approaches can further enhance model performance and practical relevance. Incorporating additional explanatory variables could improve predictive precision, such as renewable generation forecasts, cross-border flows, or demand indicators. Furthermore, model combinations or adaptive ensemble methods may help leverage complementary model strengths while maintaining computational efficiency.

About FORRS

FORRS is a leading strategy and management consultancy with a strong focus on the entire Trading Value Chain. We support clients with complex, business-critical processes in the Energy and Financial sectors.

At FORRS, we carry out projects in electricity, gas, certificates, and commodities trading. We focus on solving problems along the entire value chain in the trading sector; from front to middle to back office. Our expertise is in developing tailor-made solutions.

In addition, we offer services that go beyond consulting, as we developed the software platform GRYT. Its intelligent, cloud-based framework provides an integrated market data management system, as well as a platform to develop, standardize, and execute individual models and processes. GRYT acts as a central platform to benefit data-sensitive businesses and leverage and grow trading businesses.

Problem-focused, solution-driven. This is the difference maker. This is FORRS.

This whitepaper was developed in close collaboration with the **University of Hamburg**, under the supervision of **Prof. Dr. Alexander Szimayer**, and benefited from the valuable support of two PhD students, **Julian Sauerbier** and **Nick Frowerk**. Their expertise and contributions were instrumental in shaping the analysis, ensuring methodological rigor, and providing critical insights throughout the study.

FORRS GmbH
Dachauer Straße 63
80335 München | Germany

FORRS Office Frankfurt
Große Gallusstraße 16-18
60312 Frankfurt am Main

Phone: +49 89 38 16 45 59
Email: mailbox@forrs.de

www.linkedin.com/company/forrs

www.instagram.com/forrs_gmbh

www.forrs.de

References

[1] EPEX. "Annual trading results of 2024," Annual Trading Results of 2024 – Power Trading on EPEX SPOT reaches all-time high. (Jan. 28, 2025), [Online]. Available: https://www.epexspot.com/sites/default/files/download_center_files/ 2025 - 01 - 28 _ EPEX % 20SPOT _ Annual % 20Power % 20Trading% 20Results% 202024 _ finaldraft_0.pdf (visited on 10/20/2025).

[2] J. H. Friedman, "Greedy function approximation: A gradient boosting machine.," The Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, Oct. 2001, Publisher: Institute of Mathematical Statistics, ISSN: 0090-5364, 2168-8966. DOI: 10.1214/aos/1013203451. [Online]. Available: https://projecteuclid.org/journals/annals-of-statistics/ volume-29/issue-5/Greedy-function-approximation-A-gradient-boosting- machine/10.1214/aos/1013203451.full (visited on 02/18/2025).

[3] "FlexIndex." (Jan. 28, 2025), [Online]. Available: https://flex-power.energy/de/ flex-index/ (visited on 10/29/2025).