

## "Data Science Challenge: Predicting Energy Prices"

Day-ahead prices are the prices set for electricity delivery on the next day, determined through a daily auction process based on anticipated demand and supply. These prices provide a forecast that helps producers and consumers plan their generation and consumption, reflecting expected market conditions. Predicting these prices is crucial for energy companies and traders to optimize their operations, manage risks, and improve market efficiency.

As the goal of the upcoming BTW 2025 Data Science Challenge, you will build a model to predict hourly day-ahead energy prices for Germany on February 18, 2025 as reported by the SMARD.de information platform.

The task of creating a submission for the BTW 2025 Data Science Challenge can be partitioned into 4 phases as described below.

### Phase 1: Gathering Domain Knowledge & Data Sources

For building a model that predicts energy prices, you are required to gather not only historical data on energy prices, but also gain a strong understanding of the energy sector's dynamics. This entails collecting data for recent years on energy prices as well as demand and production capacities, weather conditions, supply metrics, and any significant events impacting energy prices. Additionally, acquiring domain knowledge is crucial for understanding the factors that influence energy markets.

You can think of your submission as a story about data discovery, where the background information sets the stage for understanding the significance and narrative of your analytical findings. It should therefore include important facts on the domain that you are exploring, for example:

- Research the energy market's structure, including how prices are set and the role of different energy sources.
- Understand regulatory policies, market reforms, and technological advancements that could impact energy prices.
- Review energy sector publications, expert analyses, and videos or other resources.

Possible data sources include energy market databases, weather archives, government publications on energy, and news articles reporting on the energy sector. Combining comprehensive data collection with indepth domain knowledge will lay a solid foundation for developing accurate predictive models and meaningful analysis.

### Phase 2: Data Cleaning & Exploratory Data Analysis (EDA)

Understanding and optimizing the quality of your data is critical for uncovering insights from your data sources and setting the stage for predictive modeling, for example:

- Prepare your data for analysis by handling missing values, outliers, and ensuring data quality.
- Use statistical summaries and visualizations to understand the basic characteristics of the data. Explore the distribution of key variables and identify patterns or anomalies.
- Investigate the relationships between different variables, especially how various factors such as weather conditions, demand, and significant events correlate with

energy prices. Utilize statistical methods like Pearson's or Spearman's rank correlation coefficient to quantify the strength and direction of these relationships. This step is crucial for identifying potential predictors for your models.

This initial analysis will not only help you understand the data better but also guide your choice of features for the predictive modeling phase. It's essential to document any insights or interesting findings during this stage, as they could be valuable for the final report and presentation.

### Phase 3: Visualization & Storytelling

Effective presentation is key to interpreting the data and predictive modeling results comprehensively, for example:

- Create visualizations that illustrate the historical trends in energy prices and highlight the key factors influencing price fluctuations.
- Use graphs to compare actual vs. predicted energy prices to evaluate the model's performance visually.
- Explain your findings by connecting the dots between visuals, forecasts, and insights, making it easier to understand how energy prices change and what that means for people and businesses.

Visualizations should not only aim to support the narrative of your findings, but also uncover insights that might not be immediately apparent from the raw data or statistical metrics alone. A well-crafted visualization can make complex data more accessible and insights more compelling.

### Phase 4: Predictive Modeling

To forecast average hourly energy prices on the target date, you may consider various modeling techniques such as:

- Linear Regression
- Time Series Analysis (e.g. ARIMA)
- Machine Learning models (e.g. Random Forest, Neural Networks)

Evaluate different models based on reasonable performance metrics (e.g. RMSE, MAE) to select the bestperforming model.