NTNU
Kunnskap for en bedre verden

Department of Computer Science

TDT4259 - Applied Data Science

Electricity Spot Prices Forecasting

Author:
Pol Fonoyet - 128703 - polf@stud.ntnu.no
Gorka Parra - 128705 - gorkap@stud.ntnu.no
Ronja Steinfurth - 128730 - ronjacst@stud.ntnu.no
Ana Barrera Novas - 128680 - anabar@stud.ntnu.no
Pol Batallé Largo - 128691 - polba@stud.ntnu.no
Daniel Alcolea Aguado - 128695 - daniealc@stud.ntnu.no
Jacob Ørmen Clements - 546272 - jacob.o.clements@ntnu.no

Autumn 2024

Table of Contents

1 Introduction and Problem Definition

This report is the result of a group assignment for the course Applied Data Science (TDT4259) at the Norwegian University of Science and Technology. Throughout the course, we have explored and applied data science concepts essential for understanding and addressing real-world problems. Our project focuses on Day-ahead Electricity Spot Price Forecasting, an issue highlighted by a research scientist from SINTEF, one of Europe's premier independent research organizations. Leveraging the provided dataset, our team's expertise, and new research insights, we aim to develop a robust analytical model to accurately forecast electricity prices in the Nord Pool market. We will now briefly introduce Nord Pool, outline the identified problem, and introduce our team, with the first two topics expanded upon in Section 2: Background.

1.1 Introduction to Nord Pool

Nord Pool is the largest electricity market in Europe, facilitating the efficient trading of electricity across multiple countries [Nor24a]. Established to enhance the coordination and optimization of electricity markets, Nord Pool plays a crucial role in balancing supply and demand in real-time [Nor24c]. The market operates in a highly dynamic environment, where prices are determined through the interaction of buyers and sellers, ensuring transparency and competitiveness [Nor24b].

In this market, participants place bids for electricity for the next 24 hours. The prices are set one day before delivery, based on the equilibrium between supply and demand. This market mechanism allows participants to plan their production or consumption schedules efficiently, ensuring that electricity needs are met at the lowest possible cost [Nor24b].

Nord Pool's role goes beyond price-setting; by facilitating a competitive and transparent environment, it ensures that electricity is produced and consumed in alignment with real-time market conditions [Nor24c]. This system is integral in reducing the likelihood of supply shortfalls or surpluses, ultimately supporting a resilient and responsive energy market [Nor24c].

1.2 Problem Definition

Our aim is to develop a model to forecast electricity spot prices 24 hours ahead in the Nord Pool market using historical spot price, production and volume data.

This problem statement lays the foundation for building predictive models that can accurately forecast future electricity prices within the Nord Pool market. These spot price predictions are essential for improving market efficiency and stability [Wer14].

Although Nord Pool facilitates electricity trading, our project focuses on leveraging data analytics to enhance forecasting accuracy. These forecasts provide substantial benefits to various stakeholders, including energy producers, traders, and large industrial consumers, by supporting informed decision-making in electricity production, trading, and consumption. This, in turn, contributes to more efficient energy use and cost reduction.

Addressing this problem involves several inherent challenges in electricity price forecasting. Electricity prices are highly volatile, influenced by factors such as weather conditions, fluctuating demand, and the availability of energy resources. To produce reliable forecasts, our model must account for these complexities by leveraging historical data effectively. Since our approach relies solely on historical data for both training and inference, ensuring data quality is imperative. Hence, we will carefully assess and preprocess the dataset to address any inconsistencies, missing values, or anomalies that could impact the model's performance.

Choosing the right forecasting model is equally important. We plan to experiment with various machine learning and statistical approaches to determine which model best captures the underlying patterns in the data. Model accuracy and stability will be evaluated using metrics relevant to time-series forecasting. Lastly, different stakeholders have unique needs and varying risk tolerances; for example, industrial consumers often prioritize price stability to manage operational costs, while traders may look to capitalize on price fluctuations [Int22]. Our forecasting model must be adaptable to accommodate the diverse objectives of these stakeholders.

1.3 Relevance of Data Science in Modern Energy Markets

This project—focused on forecasting electricity spot prices 24 hours ahead in the Nord Pool market—reflects the growing role of data science in transforming the energy sector. As the industry generates vast amounts of data, AI and machine learning are unlocking insights that drive efficiency, sustainability, and resilience [Int23]. Research organizations like SINTEF are at the forefront of using data science to tackle complex energy challenges, from optimizing resource allocation to enhancing grid stability.

Data science applications in energy markets have facilitated critical advancements, such as improved renewable energy forecasting, optimized power generation schedules, and enhanced energy trading. For instance, predictive models help grid operators manage fluctuations in renewable output, balancing intermittent supply with demand to stabilize prices [Wor23]. In a similar way, our model leverages historical price, production, and volume data to forecast electricity prices, empowering Nord Pool participants—including producers, traders, and large consumers—to make informed decisions that reduce costs and improve operational planning in a volatile market.

By aligning with these data-driven practices, our project supports the energy sector's shift toward more efficient and sustainable resource management, demonstrating how accurate price forecasting can contribute to a more resilient and adaptable energy market.

1.4 Team Description and Responsibilities

The team for this project comprises a diverse group of students from the Norwegian University of Science and Technology (NTNU) and Erasmus exchange students from the Polytechnic University of Catalonia, the University of Barcelona, the University of A Coruña, and RWTH Aachen University. Each member brings unique skills and academic backgrounds to support various aspects of the project. The table below provides an overview of each team member's expertise and their specific roles in contributing to the project's success.

Name	Background	Responsibilities
Pol Fonoyet	Exchange student from the Polytechnic University of Catalonia, pursuing a Bachelor's degree in Informatics Engineering with a major in Computing.	Preparing data, analytical research and model creation
Pol Batalle	Exchange student from the University of Barcelona, pursuing a Bachelor's degree in Telecommunication Electronic Engineering	Visualization and data analysis
Jacob Clements	As a fourth-year student in Industrial Economics and Technology Management at NTNU, Jacob specializes in Artificial Intelligence and Finance within Computer Science.	Contributes valuable business insights and creates intuitive visualizations, such as dashboards, to facilitate decision-making for stakeholders.
Ana Barrera	Exchange student from the University of A Coruña studying a Bachelor's degree in Computer Science, with a focus in Software Engineering	Business insight, project design, and video creation
Gorka Parra	Exchanage student from the *Universitat Politècnica de Catalunya*, studying a Bachelor's degree in Computer Science and focusing on artificial intelligence.	Preparing data, analytical research and model creation
Ronja Steinfurth	Exchange student from RWTH Aachen University, studying Electrical Engineering and Business Administration at Master's level	Business insight, understanding business needs and project planning
Daniel Alcolea	Exchange student from the University of Barcelona, pursuing a Bachelor's degree in Telecommunication Electronic Engineering	Visualization and data analysis

The team established an effective dynamic from the outset, prioritizing efficient collaboration, regular communication, and inclusive decision-making. We held regular meetings on campus to discuss progress and align on upcoming tasks, while Jacob, who was off-campus for the semester, remained engaged and updated through WhatsApp along with the rest of the team to facilitate real-time communication. Frequent checkpoints were established throughout the project to ensure steady progress, allowing team members to stay informed and provide input on recent developments. We also adopted a collaborative approach to each project section, with team members contributing insights and reviewing each other's work to ensure multiple perspectives and enhance the overall quality of the output. This approach contributed to a supportive and adaptable workflow, enabling us to maintain a consistent pace and produce a cohesive final result.

2 Background

Forecasting electricity spot prices is a vital tool for stakeholders in the energy sector, particularly energy producers, who need to make informed decisions on production scheduling, bidding strategies, and resource management. In this project, we aim to develop an accurate forecasting model for the Nord Pool market using historical data on demand, production, and spot prices. The model will enable stakeholders to optimize operations by predicting price fluctuations and aligning production with demand trends, ultimately enhancing their revenue and operational efficiency.

This chapter is organized as follows: First, we outline the specific business and technical objectives of the project in Section 2.1, emphasizing the needs of energy producers within the Nord Pool market. Next, we provide a literature review in Section 2.2, exploring data-driven approaches to electricity price forecasting and highlighting relevant real-world examples. In Section 2.3, we present our problem approach, outlining the strategic steps taken to develop a day-ahead forecasting model that aligns with the operational needs of energy producers in the Nord Pool market. Finally, we introduce the CRISP-DM framework and Design Thinking methodology in Section 2.4, detailing how these methodologies structure and guide the project's development, evaluation, and deployment phases.

2.1 Objectives

In our electricity spot price forecasting project, various stakeholders may benefit from accurate predictions. This section outlines the specific objectives that should be prioritized when considering energy producers as primary stakeholders. By addressing these objectives, the forecasting model will support energy producers in making informed decisions about production, bidding, and resource allocation in the Nord Pool market.

Business Objectives

The primary business objectives of this project are to support energy producers in the Nord Pool market by providing precise spot price forecasts. These forecasts enable better operational planning, revenue maximization, and strategic bidding, ultimately driving greater efficiency and profitability. The key business objectives include:

- **Align Production with Demand Trends**: Enable energy producers to optimize production schedules based on demand forecasts, allowing them to capitalize on high-demand periods to maximize revenue.
- **Develop Strategic Bidding Tactics in Day-Ahead Markets**: Use spot price forecasts to inform competitive bidding strategies in the Nord Pool day-ahead market, ensuring both profitability and market positioning.
- **Enhance Short-Term Planning and Mitigate Risks**: Provide insights into price fluctuations, allowing producers to adjust schedules and reduce exposure to unprofitable periods, thereby improving financial stability.
- **Optimize Resource Allocation for Operational Efficiency**: Improve resource allocation by minimizing overproduction or underproduction costs, contributing to overall cost efficiency and sustainability.
- **Increase Profitability from Flexible Assets**: For assets with flexible production capabilities, such as gas or hydro plants, utilize spot price trends to enable real-time adjustments and capitalize on profitable opportunities.

Technical Objectives

The technical objectives of this project focus on developing a reliable and adaptable forecasting model that meets the unique demands of the Nord Pool day-ahead market. Key technical objectives include:

- **Build an Accurate Spot Price Forecasting Model**: Develop a predictive model using historical trends in demand, production, and spot price data to capture essential relationships among these variables, aiming to deliver precise forecasts for day-ahead pricing.
- **Incorporate Temporal Patterns**: Leverage datetime information to model daily, weekly, and seasonal patterns in electricity prices, utilizing time series analysis to account for regular demand and supply fluctuations.
- **Model Volume-Price Dynamics**: Analyze `volume_demand` and `volume_production` to understand and quantify how supply and demand shifts affect pricing, ensuring the model adapts effectively to market dynamics.
- **Enable Real-Time Adaptability for Day-Ahead Markets**: Design the model to incorporate continuously updated data, allowing it to refine predictions in response to new volume and price inputs, thus supporting agile decision-making in a fast-paced market.
- **Ensure Robust Evaluation and Performance Monitoring**: Use reliable evaluation metrics (e.g., Mean Absolute Error, Root Mean Squared Error) to track and assess model performance, ensuring accuracy.

2.2 Literature Review: Data-Driven Approaches to Electricity Price Forecasting

Accurate forecasting of electricity prices is essential in energy markets for managing resources effectively and reducing operational costs. Data-driven methods have demonstrated significant potential in meeting similar forecasting challenges.

The study "Data-Driven Techniques for Short-Term Electricity Price Forecasting" explores methods specifically aimed at predicting electricity prices over short time frames. It highlights the importance of capturing non-linear relationships in price data, a factor that supports our use of machine learning models, such as Random Forest, which are well-suited for handling such complexities [MDP23]. This research reinforces our model choice by demonstrating the effectiveness of machine learning techniques in single-step, day-ahead price forecasting.

Another relevant review, "Energy Price Modelling: A Comparative Evaluation of Four Generations of Models," assesses the evolution of energy price models, discussing both statistical and machine learning approaches. The study emphasizes that advanced machine learning models are particularly effective for short-term predictions due to their flexibility in learning from historical data [arX24]. This supports our goal of using a data-driven model to capture the dynamic patterns in Nord Pool electricity prices accurately.

In a real-world example, forecasting models have been successfully applied in the Australian Energy Market, where accurate electricity price predictions have allowed producers to optimize operations and enhance profitability by anticipating demand and adjusting supply accordingly. This example illustrates the broader impact that reliable price forecasting can have on operational efficiency and resource management, underscoring the practical value of our project in supporting data-driven decision-making [Aus23].

These studies and applications validate our approach to electricity price forecasting by highlighting the suitability of machine learning models in achieving accurate, short-term predictions. Our project aims to leverage these insights to develop a robust model that enhances energy producers' decision-making capabilities in the Nord Pool market.

2.3 Problem Approach

The primary aim of this project is to develop a reliable day-ahead electricity spot price forecasting model tailored for the Nord Pool market. Our approach centers on building a data-driven solution that enables for example energy producers to make well-informed decisions, particularly in production scheduling and strategic bidding. By analyzing historical data on demand, production, and pricing, we seek to capture the essential patterns and trends that influence daily price fluctuations.

This approach follows a structured series of steps, starting with data collection and feature selection, ensuring that relevant temporal and seasonal patterns are integrated into the model. We then apply forecasting techniques that align with the unique requirements of the day-ahead market, focusing on accuracy and practical applicability for industry use. Each stage of the approach is designed to provide clear, actionable insights that can enhance energy producers' operational planning and financial outcomes.

Ultimately, our problem approach aims to deliver a robust forecasting model that supports strategic and operational goals, allowing stakeholders to navigate the complexities of the Nord Pool market more effectively.

2.4 Project Management Strategy

This chapter outlines the CRISP-DM framework, a widely adopted methodology for managing data mining projects. The structure of this report is organized around the distinct phases of the CRISP-DM framework, highlighting the specific tasks and objectives within each phase. Additionally, Design Thinking serves as an important complement to CRISP-DM, enhancing the project management strategy by emphasizing user-centered approaches and iterative problem-solving throughout the data mining process.

2.4.1 Crisp-DM Framework

The CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology, developed in 1996, is a well-established framework for guiding data mining and data science projects. As an open standard, it provides a structured, field-proven approach that has been widely adopted across industries. CRISP-DM serves as a comprehensive process model, outlining the typical phases of a data project, detailing each phase's specific tasks, and clarifying the relationships between them, thus offering a clear overview of the entire data mining life cycle.

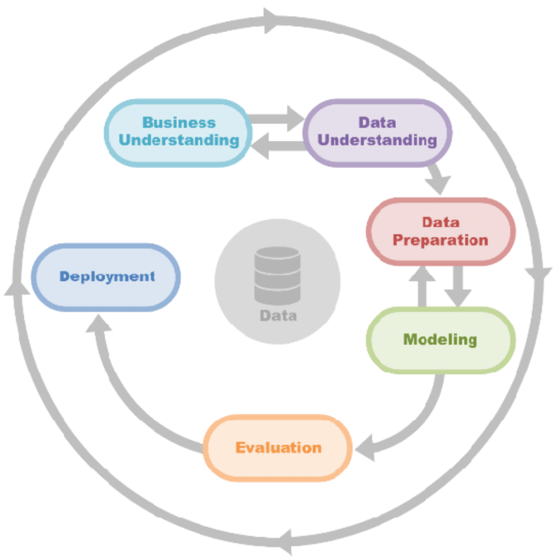


Figure 1: CRISP-DM Framework [TAH20]

This methodology divides the data mining process into six essential phases: business understanding, data understanding, data preparation, modeling, evaluation, and deployment, as illustrated in Figure 1. Each phase serves a distinct purpose and supports iterative improvement, enabling agile transitions that adapt to evolving project requirements. The CRISP-DM structure not only ensures alignment between technical efforts and business objectives but also facilitates effective project management by clearly defining each step's role in producing actionable results. A comprehensive overview of each phase is presented in the following sections, drawing on insights from [She00].

Phase 1: Business Understanding

The Business Understanding phase sets the project's foundation by defining objectives from a business perspective and aligning them with strategic goals. Key steps involve setting objectives, assessing the current situation, and outlining a project plan to ensure relevant data aligns with business needs.

Phase 2: Data Understanding

Data Understanding begins with data collection and familiarization, identifying quality issues, and uncovering patterns. Steps include gathering relevant data, describing characteristics, exploring patterns, and verifying quality to ensure reliable analysis.

Phase 3: Data Preparation

Data Preparation transitions raw data to a final dataset suitable for modeling. Key tasks include selecting relevant data, cleansing inaccuracies, constructing new variables, integrating sources, and formatting for compatibility with modeling tools.

Phase 4: Modeling

In the Modeling phase, various techniques are applied, and parameters optimized. Key steps include selecting algorithms, designing tests, building models, and assessing effectiveness, generating actionable insights from the data.

Phase 5: Evaluation

The Evaluation phase critically assesses the model's performance, ensuring alignment with business objectives. Key actions include result evaluation, process review, and determining next steps for decision-making.

Phase 6: Deployment

Deployment translates insights into actionable formats, from reports to integrated "live" models. Key actions involve planning deployment, establishing monitoring, preparing reports, and conducting a project review to support practical application.

2.4.2 Overview of CRISP-DM Framework Integration

This section provides a concise summary of how each phase of the CRISP-DM methodology is applied across different sections of this report. Each phase contributes unique steps that are critical for addressing the project's objectives, guiding the analysis, and ensuring the model's practical application.

In Section 1.2 and 2.1, the Business Understanding phase is implemented, setting the foundation for defining project goals and aligning them with strategic objectives. The Data Understanding phase, outlined in Section 3.1 and 3.2, focuses on initial data exploration and quality assessment, laying the groundwork for informed analysis.

Section 3.3 documents the Data Preparation phase, detailing the transformation, cleaning, and structuring of data necessary for robust modeling. Modeling is covered in Section 3.4 and 3.5, where various techniques are applied, and their parameters calibrated to ensure optimal outcomes. In Section 4, the Evaluation phase rigorously tests the model's performance and ensures alignment with business requirements.

Finally, the Deployment phase is addressed in Section 5 and 6, highlighting how the model's insights are integrated into actionable formats to support decision-making and long-term project goals.

Each of these sections illustrates how the phases of the CRISP-DM framework contribute to achieving a structured, actionable data science project.

2.4.3 Design Thinking

Design Thinking, though defined in various ways, consistently highlights core elements. Some perspectives emphasize its focus on human-centered, needs-based problem-solving, while others underscore its capacity to drive entrepreneurial innovation. For example, [BU16] describes Design Thinking as a process that starts with human needs and harnesses technology to create value for both customers and businesses. Rather than following traditional "analytical" management methods, Design Thinking complements the structured CRISP-DM framework by encouraging creative, customer-focused solutions within data-driven projects.

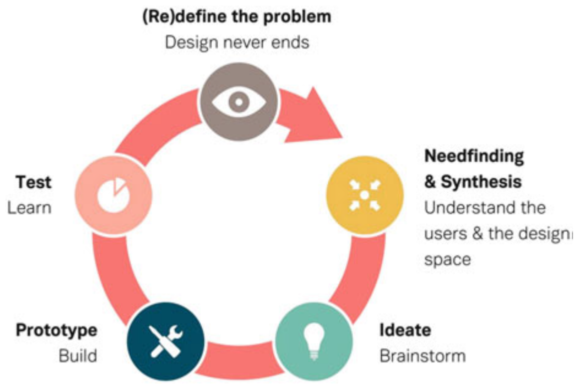


Figure 2: Design Thinking Micro-Process [BU16]

[BU16] describes the Design Thinking micro-process as an iterative cycle of six steps, as displayed in Figure 2:

1. Define the Problem: Formulate a focused yet open-ended challenge question to frame the project goals.
2. Needfinding and Synthesis: Investigate and interpret customer needs, capturing both obvious and hidden insights through research and direct interactions.
3. Ideate: Generate diverse ideas in response to the identified needs, encouraging innovative thinking grounded in real requirements.
4. Prototype: Develop tangible prototypes to explore solution possibilities, varying in complexity based on the project stage.
5. Test: Evaluate prototypes with actual users in real-world contexts, gathering critical feedback.

6. Learn: Reflect on testing outcomes to verify if the solution addresses the problem; adjust or redefine the challenge if needed, leading to new cycles.

This flexible, user-centered approach emphasizes continual refinement and responsiveness to insights throughout the process.

2.4.4 Integration of Design Thinking

The Design Thinking process guided our project in a way that kept us focused on the needs of the users and helped us adapt continuously based on new insights. We started by defining a clear problem, which was centered on the needs of energy producers in forecasting electricity prices. Each step of Design Thinking allowed us to refine our approach and make sure our solution stayed relevant to this goal.

In the Needfinding and Synthesis stage, we learned about specific and hidden needs by engaging with stakeholders and understanding their challenges in forecasting. This research gave us a strong basis to move into the Ideation phase, where we developed different ideas that could meet these needs.

During Prototyping, we turned our ideas into real forecasting models to test various methods and configurations. Each model was tested and improved to make it more accurate and practical. In the Testing phase, we looked at how well each model worked and noting strengths and weaknesses. This feedback guided us in the Learning phase, where we reviewed our results and identified areas to improve, sometimes returning to earlier steps as needed.

The Design Thinking process allowed us to learn and adjust our approach continuously. Each test phase brought new insights, leading us to revisit and refine earlier steps to ensure we were addressing the problem effectively. By combining Design Thinking with the CRISP-DM framework, we increased the flexibility of our project and created a user-focused, collaborative approach to developing our solution.

3 Method and analysis

This part outlines the methodologies and analytical approaches employed in our research. It covers data visualization, dataset characterization, data preprocessing procedures, analytical strategies, and machine learning techniques.

3.1 Dataset Description

3.1.1 Features and Attributes

The dataset includes four key features: datetime_utc, volume_demand, volume_production, and spot_price. The objective is to predict future electricity spot prices in the Nord Pool market using historical data on spot prices and volumes.

Datetime UTC

The datetime_utc feature provides the point in time of a single observation in the format YYYY-MM-DD HH:MM:SS+00:00. All datetimes are in UTC and are registered at the beginning of an hour. The first datetime is 2015-12-31 23:00:00+00:00, and the last is 2018-09-13 02:00:00+00:00, which adds up to 23,666 rows of data in total. As for the attribute type, although time itself is a continuous value, it is a discrete attribute in this dataset. This is due to the discrete nature of the measurements, where each observation pertains to a specific hour.

Volume Demand

The volume_demand is a floating point number representing the total electricity demand (in megawatts) at a specific time. This feature records the amount of electricity consumers require from the grid and can take any positive real number. As a continuous attribute, volume_demand can vary greatly depending on the time of day, season, and other factors influencing electricity consumption.

Volume Production

Similar to volume_demand, volume_production is a floating point number that indicates the total electricity produced (in megawatts) at a particular timestamp. This feature captures the supply side of the electricity market, detailing how much electricity is generated and fed into the grid. It is a continuous attribute, capable of taking any positive real value, and varies with production capacity, fuel availability, and other production factors.

Spot Price

The spot_price, which is the target variable for prediction, is a floating point number representing the price of electricity (in euros per megawatt-hour) at the specific timestamp. It reflects the real-time cost of electricity in the market and is influenced by both demand and supply conditions. Being a continuous attribute, spot_price can assume any non-negative real value, indicating the dynamic nature of electricity pricing based on market conditions.

3.1.2 Sources

The dataset used for this project originates from Nord Pool, the leading electricity market in Europe. Nord Pool provides comprehensive data on electricity spot prices and market volumes, which allows us to analyze trends and patterns within the energy market. The data covers multiple years, offering an extensive look at fluctuations in electricity prices and volumes, which is critical for forecasting and model training.

In terms of volume data, Nord Pool records both production and demand volumes on an hourly basis. Each observation represents a sum of electricity demand or production at the start of each hour across various regions participating in the Nord Pool market. These records reflect aggregate data from multiple suppliers and consumers, ensuring a broad view of market dynamics but without individual participant details.

Additionally, Nord Pool's datasets include precise timestamps for each recorded entry, aligning with UTC and timestamped at the beginning of every hour. The data structure helps us develop accurate time-series models by ensuring that every hour is accounted for without any missing entries.

3.2 Data Analysis Methods and Tools

This part outlines the methodologies and tools applied to examine the Nord Pool dataset, detailing their relevance to our research objectives. Here, we focus on identifying the data characteristics that could best support the predictive aspects of our study. Additionally, insights from this analysis informed our feature selection process, guiding us in choosing the most impactful variables for our objectives. We include descriptive statistics along with visualizations to showcase key findings from our analysis.

Seasonal and Trend decomposition using Loess

Understanding seasonality and trends is crucial in any time series analysis. Seasonality refers to a recurring pattern at a specific interval, while the trend represents a long-term upward or downward movement in the data. These elements are key in predicting future values based on historical data.

When a trend is present, it signals a persistent change that is likely to continue, while seasonality allows for the prediction of values based on the time elapsed since previous observations.

To analyze the seasonal and trend components in our time series, we employ the STL (Seasonal-Trend decomposition using LOESS) method, as developed by [Cle+90]. STL is widely used in time series analysis for its ability to decompose a series into three components: seasonality, trend, and residuals. By summing these components, we can reconstruct the original time series, making STL an effective tool for capturing and analyzing temporal patterns. This decomposition method is recognized for its robustness and versatility in handling complex seasonality, making it suitable for our dataset in the Nord Pool market [Cle+90]. Figures 3 and 4 illustrate STL plots for the Nord Pool dataset, with Figure 3 covering the entire spot price history and Figure 4 focusing on the initial 300 hours.

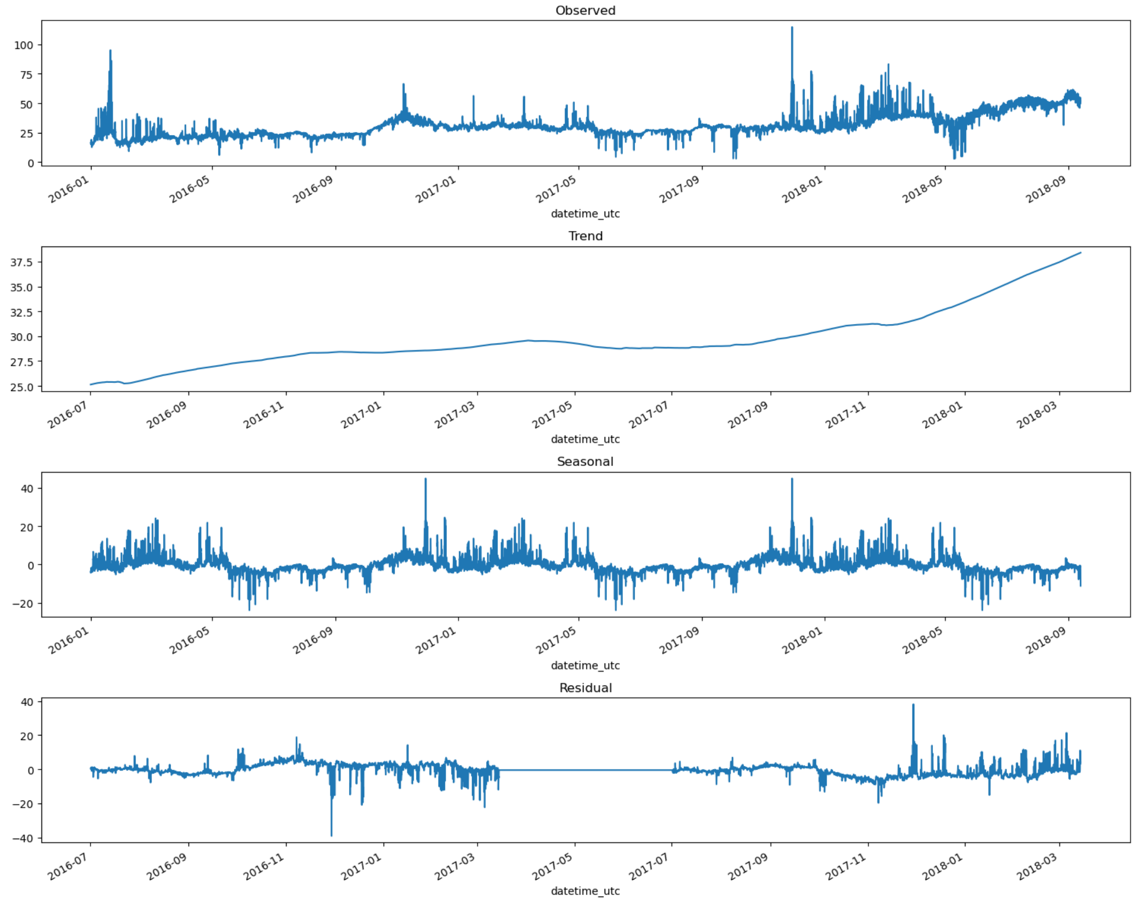


Figure 3: STL decomposition applied to spot price series for all hours

We observe a pronounced trend and seasonality within the dataset. The trend component shows a consistent upward movement throughout the period, indicating a general increase in spot prices over time. Although seasonality is evident, its frequency is too high for a qualitative analysis in figure 3. However, the STL decomposition in figure 4 provides a clearer representation:

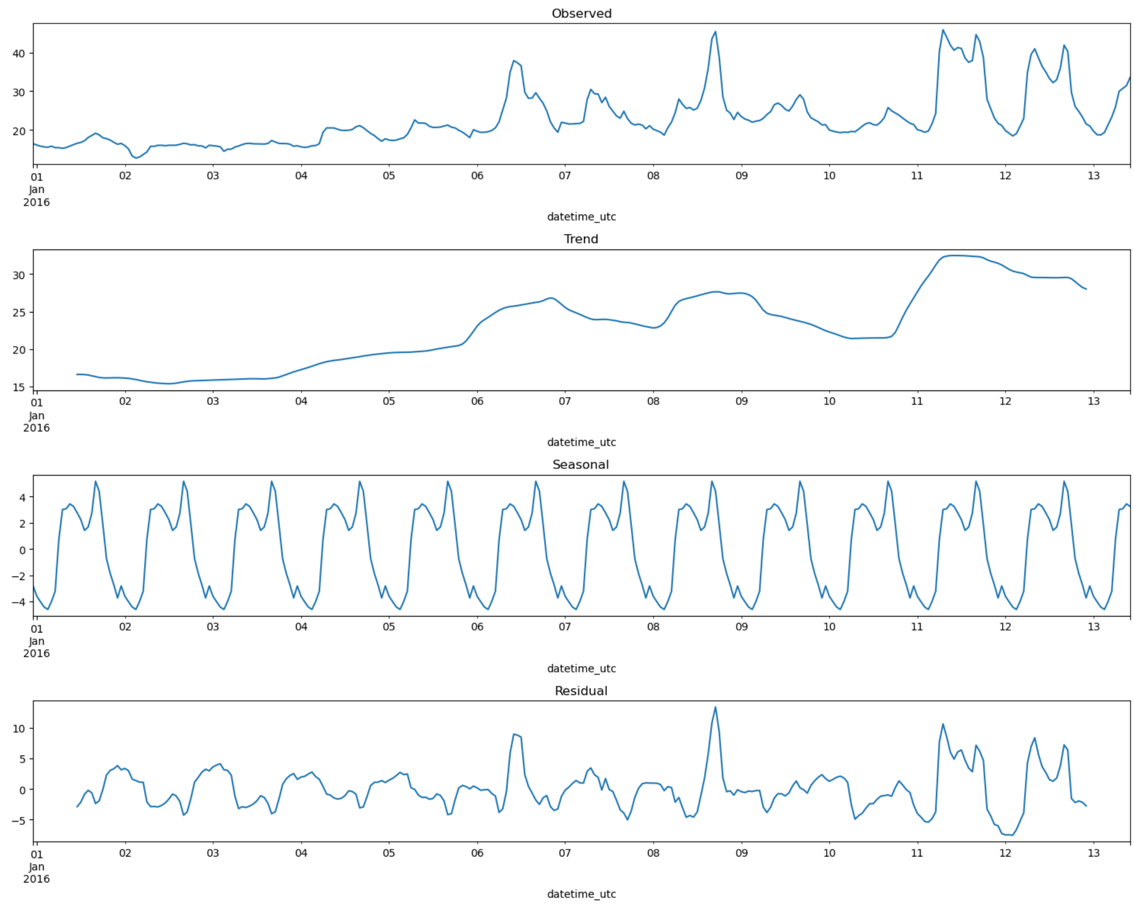


Figure 4: STL decomposition applied to the spot price series for the hours 0 to 300

From Figure 4, we observe a strong presence of intra-daily seasonality in the spot price. This makes intuitive sense, as energy demand varies with human activity. Since less energy is needed during the night, intra-daily seasonality is expected. We also hypothesized that spot prices could vary depending on the day of the week and the month. This hypothesis aligns with our understanding that energy demand—and thus spot prices—fluctuates with daily and monthly patterns in human activity. The analysis of the following figure supports these hypotheses, revealing significant variations across different days and months.

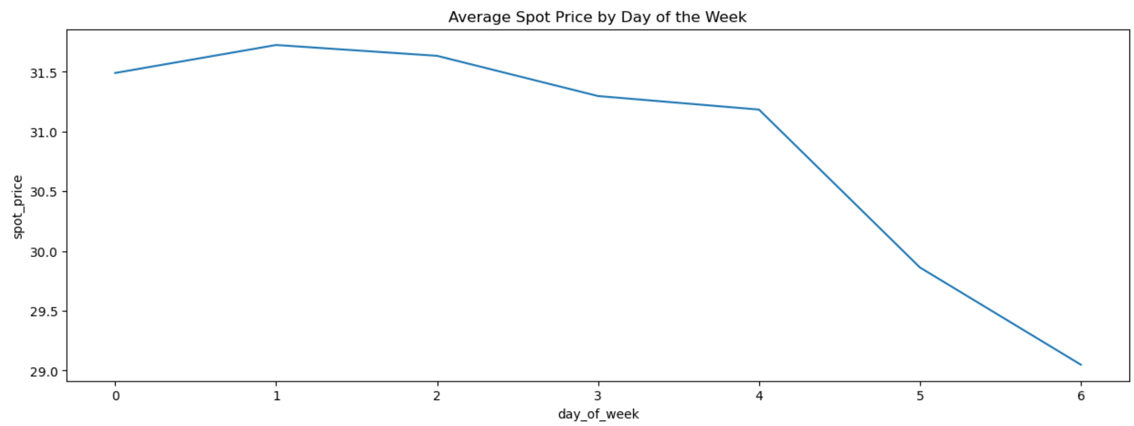


Figure 5: Average spot price by day of the week

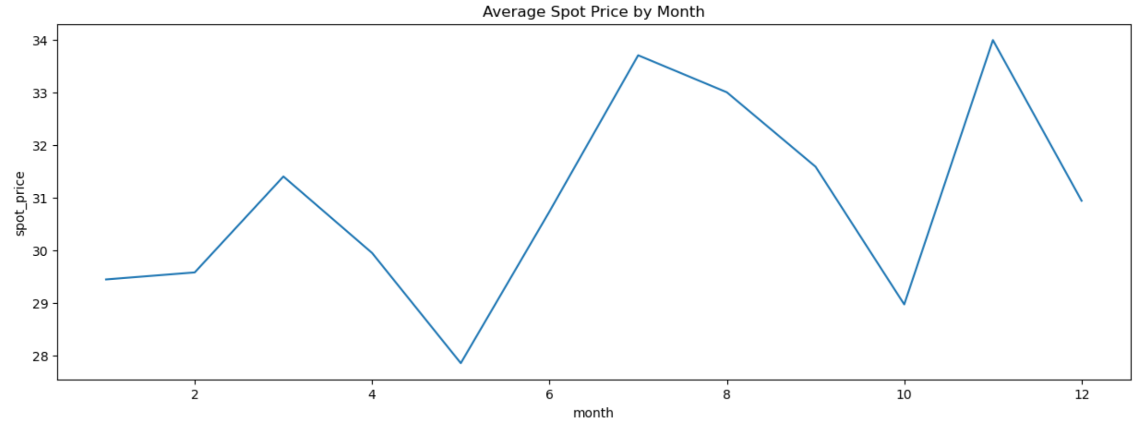
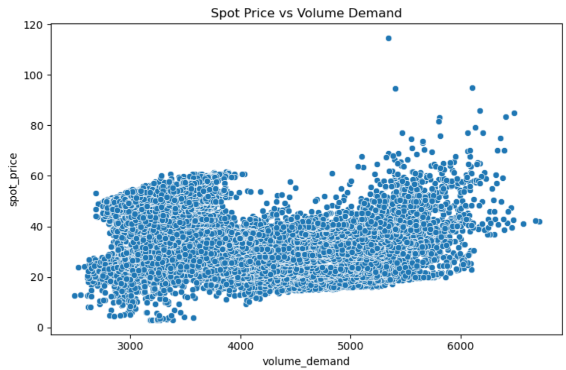


Figure 6: Average spot price by month

At the end of the week, we observe a tendency for spot prices to be lower. This trend can be attributed, in part, to decreased industrial electricity consumption. Monthly patterns may also be linked to temperature fluctuations, which significantly impact the demand for heating or cooling. We hypothesized that energy production and demand would influence its price. By creating correlation plots, we observed a slight positive correlation, indicating that when both consumption and production are higher, prices tend to increase. Additionally, when analyzing the correlation between the differences and the ratio of consumption and production with price, we found a similar slight positive correlation. This further supports our initial hypothesis.



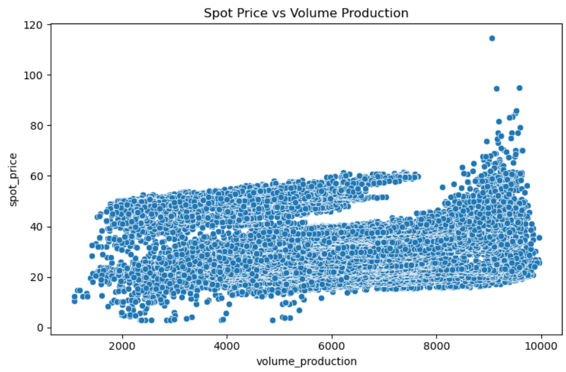
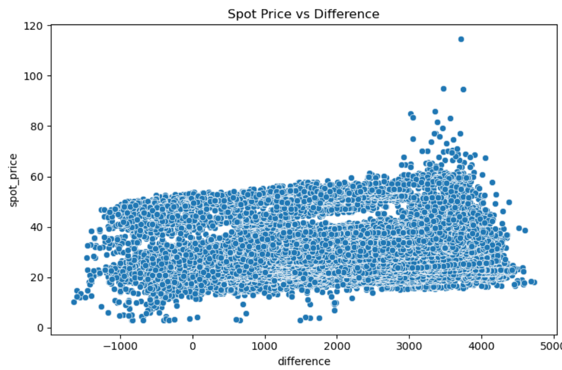


Figure 7: Demand and Production correlation plots over the entire duration of the dataset



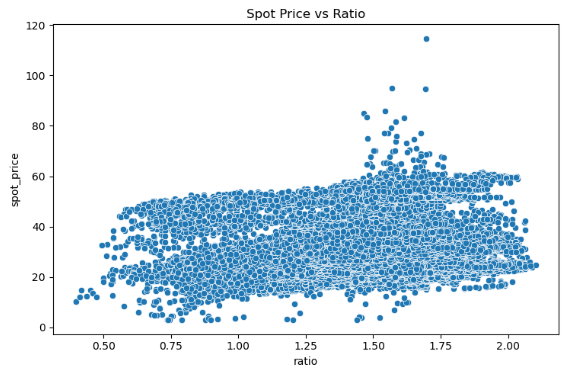


Figure 8: Difference and Ratio correlation plots over the entire duration of the dataset

Autocorrelation

Autocorrelation is a key metric in time series analysis, particularly useful for understanding seasonality and trends. It measures the correlation between a time series and a lagged version of itself over successive time intervals. This helps determine the relationship between past and future values of the series.

If X_t represents the time series at time t, and $\mu_t = E(X_t)$ is the expected value of X_t:

1. **Covariance** between X_t and X_{t+h} is given by:

$$\gamma_X(t+h, t) = \mathrm{Cov}(X_{t+h}, X_t) = E((X_{t+h} - \mu_{t+h})(X_t - \mu_t)) \tag{1}$$

2. **Autocorrelation** for lag h is defined as:

$$\rho_X(h) = \mathrm{Corr}(X_{t+h}, X_t) = \frac{\gamma_X(h)}{\gamma_X(0)} \tag{2}$$

where $\gamma_X(0)$ is the variance of X_t.

Autocorrelation provides insight into how past values of a time series influence future values. For data with a strong trend, values at nearby time points are likely to be similar, resulting in high positive autocorrelation at small lags. If the data exhibits seasonality, autocorrelation will show a repeating pattern at lags corresponding to the seasonal period. For example, if data has a yearly seasonality, autocorrelation will be higher at lags of 12 months, 24 months, etc.

High positive autocorrelation indicates that high values tend to follow high values, and low values follow low values, which is typical for trending data. High autocorrelation at specific lags suggests seasonality, as similar patterns repeat at these intervals.

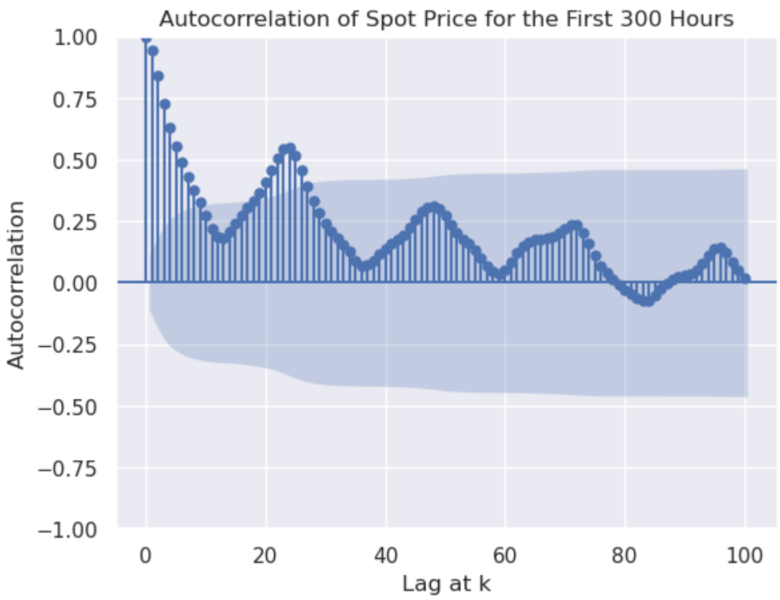


Figure 9: Autocorrelation of the spot price from hour 0 to 300

In Figure 9, it is evident that autocorrelation diminishes for time points further back, with peaks at intervals corresponding to the seasonal frequency. This behavior implies the presence of both trend and seasonality in the observed data, supporting the hypothesis of recurring patterns in the spot price series. Moreover, the analysis reveals a notable correlation between spot price values and their historical counterparts, which suggests that adding lag features could improve the model's capacity to capture these temporal dependencies. This understanding provides a basis for determining an effective range for lag features. The shaded blue area in the plot highlights the statistically significant correlations, indicating that these observed relationships are unlikely due to random noise. Autocorrelation appears to lose significance after roughly 30 time points, offering practical insights for designing our feature engineering strategy, as detailed in the next section.

Correlation Analysis

To better understand the relationships between key variables—demand, production, and spot prices—we conducted a correlation analysis. This analysis helps identify both strong and weak associations between these variables, providing insights into how fluctuations in demand and production volumes might impact spot prices. A heatmap was used to visually represent these relationships.

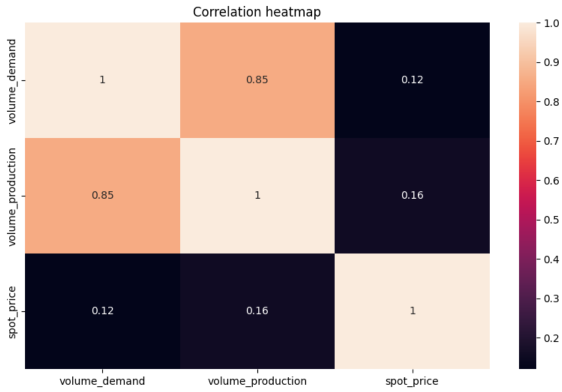


Figure 10: Correlation heat map

In the heatmap shown in Figure 10, color intensity represents the strength of correlation values, ranging from -1 to 1. A higher intensity indicates a stronger correlation, with the color bar on the right serving as a reference. We observed a high positive correlation (0.85) between volume demand and volume production, suggesting that as demand volume rises, production volume also tends to increase. This is an expected relationship, as production is often adjusted to meet demand in electricity markets.

However, the correlations between spot price and the other variables (volume demand and volume production) are much weaker. The correlation between volume demand and spot price is 0.12, indicating a very weak positive relationship. Similarly, the correlation between volume production and spot price is 0.16, also weak. These findings imply that neither demand nor production volume alone has a substantial impact on the spot price, likely due to the influence of other market factors such as external energy sources, weather conditions, and regulatory policies.

This correlation analysis supports our understanding that, while demand and production volumes are related, their direct effect on spot prices may be limited. This insight will guide our feature selection, highlighting the need to explore additional variables and complex interactions to improve the predictive power of our model.

3.3 Data Preprocessing

Feature Engineering

To enhance the predictive utility of the time variable, we decomposed it into five distinct components: day of the week, day of the year, month, hour, and year. This breakdown captures seasonal patterns at multiple granularities, enabling the model to recognize recurring trends that operate on hourly, daily, weekly, monthly, and yearly cycles. This step was based on prior analysis confirming significant seasonal behaviors within the dataset, as well as the weak correlations observed between `volume_demand`, `volume_production`, and `spot_price` in the correlation analysis (see Section 3.2). These findings guided us to prioritize time-based features over demand and production alone for capturing price variations more effectively.

In a preliminary phase of the project, we explored the addition of lag features to capture the autocorrelation observed in the data. While statistically significant autocorrelation patterns suggested that lagged features might be useful, we found that incorporating them did not notably improve model performance. Consequently, these lag features were excluded from the final model.

Target Variable Definition

Given our objective of forecasting the spot price 24 hours into the future for day-ahead market planning, we created a target variable, `spot_price_tomorrow`. This was achieved by shifting the `spot_price` column forward by 24 hours, aligning the price value from the next day with the relevant feature data from the previous day. This setup ensured that each row in the dataset contained the necessary information for one-day-ahead price prediction, directly supporting operational decision-making needs. After this transformation, the final rows, which contained NaN values for the target variable due to the shift, were removed to maintain data consistency.

Data Splitting: Training, Validation, and Test Sets

For model evaluation and tuning, we split the dataset into training, validation, and test sets. The validation set covers the final 72 hours (equivalent to three days) of data, while the test set is restricted to the last 24 hours. This temporal split respects the time series nature of the data, ensuring that future values are only used for testing purposes, thereby preventing data leakage. This approach maintains the integrity of model evaluation, as the model does not see future data during training, aligning with best practices in time-series forecasting. For final evaluation, we used Mean Squared Error (MSE) as the metric to assess prediction accuracy.

Preprocessing

The `datetime_utc` column, containing timestamp information, was converted to a datetime format using `pandas.to_datetime()`, allowing for easy extraction of time-based features such as year, month, day, and hour. After extracting these features, we removed the `datetime_utc` and `date` columns, as they no longer provided additional information relevant to our analysis, and datetime type did not work with our chosen ML model.

Data Cleaning

As part of data cleaning, we verified that no missing values were present across any columns, including `volume_demand`, `volume_production`, and `spot_price`. This lack of missing values ensured the reliability of our dataset for modeling without needing to impute or drop rows.

Through these preprocessing steps, we prepared a clean and structured dataset that focuses on essential features while capturing temporal patterns relevant to forecasting electricity prices.

3.4 Machine Learning Methods

Model Choice and Motivation

We selected Random Forest regression as our forecasting model due to its robustness and strong performance in capturing complex, non-linear relationships in time series data [Bro23b]. Its quick implementation using the scikit-learn library made it a practical choice for our one-step-ahead electricity price forecasting.

If we were to consider multi-step forecasting, Random Forest could be used in a recursive manner, where each predicted value is fed back into the model as an input for the next time step. However, this approach tends to lead to error accumulation, as early prediction inaccuracies propagate across subsequent forecasts [Bro23a]. Thus, if our objective were to shift towards multi-step predictions, we would need to reconsider the model, potentially opting for sequential models like Long Short-Term Memory (LSTM) networks that are designed to handle temporal dependencies effectively [ZPZ21]. Nevertheless, in this project we focus on one-step-ahead forecasting, which makes Random Forest a suitable choice.

We initially developed a baseline model based on recent historical values to provide a comparison and assess the predictive power of the Random Forest model. The results section presents a comparative analysis of these methods, offering an empirical justification for selecting Random

Forest.

Baseline: Naive Forecasting

A naive forecasting approach for predicting the next day's spot price simply replicates the most recent day's data, assuming persistence in the price pattern. In our context, this baseline model predicts the spot price for the next day based on the previous day's price, representing the most straightforward prediction technique. Concretely, if $\hat{y}_i$ is the prediction and y_i is the actual spot price at time i:

$$\hat{y}_i = y_{i-24}$$

This baseline model serves as a reference for assessing the predictive power of more complex models and is evaluated in the results section.

Random Forest

Random Forest is an ensemble model composed of multiple decision trees, where each tree is trained on a random subset of the data. The model's final prediction is made by averaging the predictions from all individual trees, which reduces variance and improves robustness. For brevity, we include a condensed mathematical description of Random Forest. The model prediction is calculated by averaging the predictions from T trees:

$$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} f_t(x)$$

where $\hat{y}$ is the final prediction and $f_t(x)$ represents the prediction from the t-th tree.

The objective of the Random Forest algorithm is to minimize the mean squared error, defined as:

$$L(\phi) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

where N is the number of observations, y_i is the actual value, and $\hat{y}_i$ is the predicted value.

In summary, Random Forest combines multiple weak learners (decision trees) to create a strong learner. Its ensemble structure helps it generalize well, making it suitable for our forecasting task, as it captures complex patterns and minimizes overfitting across varying time intervals. For a more detailed treatment of Random Forests, including theoretical underpinnings and applications, see [Bre01; HTF09].

3.5 Model Evaluation and Metrics

Metrics

To evaluate our model's performance, we use Mean Squared Error (MSE) as the primary metric. MSE is particularly suited to this dataset as it penalizes larger errors more heavily, encouraging a model that minimizes significant deviations in forecasted values.

$$\text{MSE} = \tfrac{1}{n}\textstyle\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Train, Validation, and Test Split

Our dataset was divided into a training set, a validation set, and a test set. The test set comprises the last 24 hours of data, and the validation set consists of the preceding 72 hours, which we further split using a sliding window approach for model tuning. This approach allows us to evaluate the model's performance over three sequential 24-hour windows within the validation period.

By structuring our data in this way, we avoided data leakage, ensuring that only past data is used to predict future values. This separation preserves the temporal integrity of the time series, which is essential for realistic model evaluation.

Hyperparameter Tuning with Optuna

We used the Optuna library to perform hyperparameter tuning, optimizing for MSE on the validation windows. Key parameters tuned through Optuna include:

- `n_estimators`: Number of trees in the Random Forest.
- `max_depth`: Maximum depth of each tree.
- `min_samples_split`: Minimum number of samples required to split a node.
- `min_samples_leaf`: Minimum number of samples required to be at a leaf node.

Optuna's optimization process allowed us to find a combination of hyperparameters that balanced model complexity and performance effectively, using MSE as the tuning criterion.

In summary, this evaluation framework—comprising the use of MSE, a time-based sliding window validation split, and systematic hyperparameter tuning—ensures our model's robustness and suitability for one-step-ahead electricity price forecasting while safeguarding against data leakage.

4 Evaluation and interpretation

4.1 Results of Predictions

Our objective was to develop a machine learning model capable of accurately forecasting electricity spot prices one day in advance, to support real-time decision-making in the Nord Pool electricity market. Using a Random Forest model, we predicted the spot prices over the last 4 days of the training set, the 3-day validation set, and the 1-day test set, demonstrating the model's ability to track short-term price fluctuations.

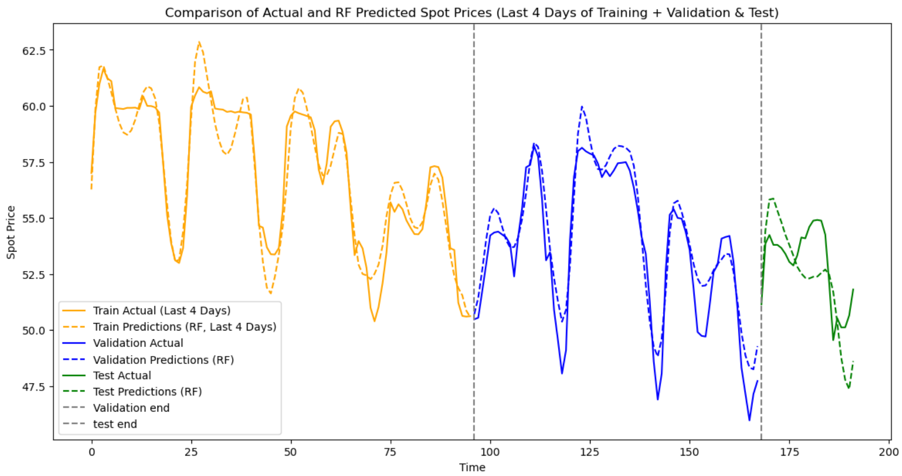


Figure 11: Comparison of Actual and Random Forest Predicted Spot Prices (Last 4 Days of Training + Validation and Test)

Figure 11 presents a comparison of actual spot prices against the model's predictions for the training (orange lines), validation (blue lines), and test periods (green lines). During the training phase, the model closely follows the historical price fluctuations, demonstrating effective learning of recent data patterns. In the validation period, the model's predictions continue to align well with actual spot prices, capturing key trends and intra-daily variations.

In the test set, which represents a completely unseen 24-hour period, the model shows a good alignment with actual prices but occasionally overshoots certain peaks and valleys. This slight overestimation suggests room for refinement, as the model's current structure may slightly amplify price movements. Addressing this tendency could further enhance predictive accuracy, particularly for periods of extreme price changes.

The day-ahead forecasting performance shown here illustrates the model's potential value for market participants by accurately predicting general trends while leaving some areas for improvement. By refining the model, potentially with additional exogenous factors such as weather conditions or economic indicators, predictive precision during volatile periods could be improved.

Overall, these results support the achievement of our main project objective, demonstrating that our model can provide reliable day-ahead forecasts to assist in operational planning and market strategy within the Nord Pool environment.

4.2 Error Analysis

As discussed in Section 3.4, our selection of the Random Forest model was motivated by its ability to capture complex, non-linear relationships in one-step-ahead forecasting. Given our objective of forecasting electricity prices for the next day, Random Forest serves as a robust choice for this short-term horizon, where the model's lack of inherent temporal dependency handling does not pose a significant limitation.

To evaluate the model's performance, we used Mean Squared Error (MSE) as our primary metric. MSE is valuable in the context of electricity price forecasting, as it penalizes larger errors more heavily, making it suitable for applications where substantial prediction deviations carry significant economic implications. In the context of Random Forest, MSE helps reveal potential overfitting or underfitting by comparing training, validation, and test errors.

To understand the behavior of the model over extended forecast horizons, we also examined the cumulative MSE across a week of predictions, as illustrated in Figure 12. This analysis, while outside the main focus of our project, provides insights into the limitations of Random Forest in multi-step forecasting.

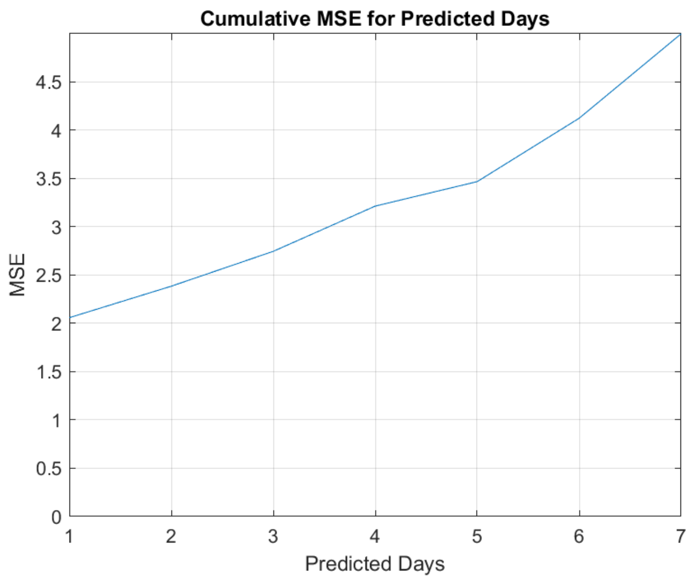


Figure 12: Predicted MSE for the next week

As shown in Figure 12, the MSE progressively increases as the forecast horizon extends. This is expected, as error accumulation is a known issue with recursive forecasting methods, where predictions for subsequent time steps are based on earlier predictions, compounding inaccuracies over time. This limitation aligns with the insights presented in Section 3.4, where we noted that Random Forest models are not well-suited for multi-step forecasting due to the propagation of error across predictions. A sequential model, like LSTM, would handle temporal dependencies more effectively in this context.

In summary, while the Random Forest model demonstrates reliable accuracy for day-ahead forecasting, its limitations in long-term forecasting are evident in the cumulative error trend. This analysis reinforces that Random Forest is well-suited for our primary goal of one-step-ahead predictions, and that caution should be exercised if applying this model for longer forecast horizons.

4.3 Feature Importance

The following chart illustrates the **feature importance** scores as calculated by the `RandomForestRegressor` model. In Random Forest, feature importance is determined by measuring how frequently a feature is used in decision nodes across all trees in the model and the extent to which it reduces prediction error when splitting the data at those nodes [Bre01]. This metric helps us understand which features contribute most to the model's predictive accuracy.

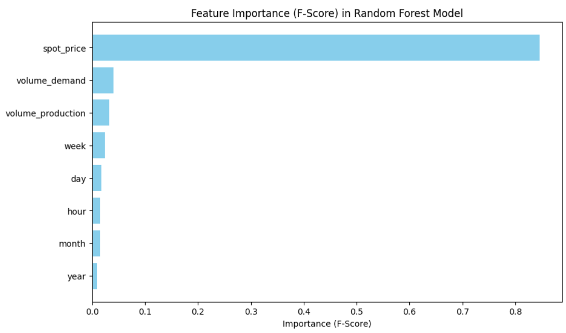


Figure 13: Feature Importance in the Random Forest Model

The feature importance plot reveals that `spot_price` is by far the most influential predictor for future spot prices, indicating that the current price is a strong indicator of short-term price movements. Other features, such as `volume_demand` and `volume_production`, also contribute to the model but with substantially lower importance.

Interestingly, the temporal features (e.g., `year`, `month`, `week`, `day`, `hour`) have minimal importance, suggesting that the specific timing within the year or week is less crucial in this model compared to direct price and volume indicators. This insight implies that the model could potentially be simplified by omitting low-importance temporal features, likely without significant accuracy loss. Such simplifications could reduce model complexity and improve interpretability, while maintaining performance.

In summary, this feature importance analysis suggests that the model's predictive capability is primarily driven by the most recent `spot_price`, which aligns with our understanding of short-term price dependencies in electricity markets. Monitoring changes in `spot_price` closely could therefore be essential for improving short-term forecasting accuracy, as the model relies heavily on this feature.

4.4 Benchmarking Random Forest with Baseline Model

To assess the effectiveness of the Random Forest (RF) model, we compared its performance against a baseline model that uses recent historical prices for prediction. This comparison helps highlight the added value of the RF model in capturing patterns and dynamics in the data that are not fully represented by simple historical methods.

4.4.1 Comparison of Predictions on Test Set

Figure 14 shows a zoomed-in view of the RF model and the baseline model predictions compared to the actual spot prices over a single day in the test set. Here, we can see that the RF model follows the actual spot price more closely than the baseline model, which often lags behind due to its reliance on prior values. This alignment with actual values indicates that the RF model successfully captures short-term fluctuations better than the baseline.

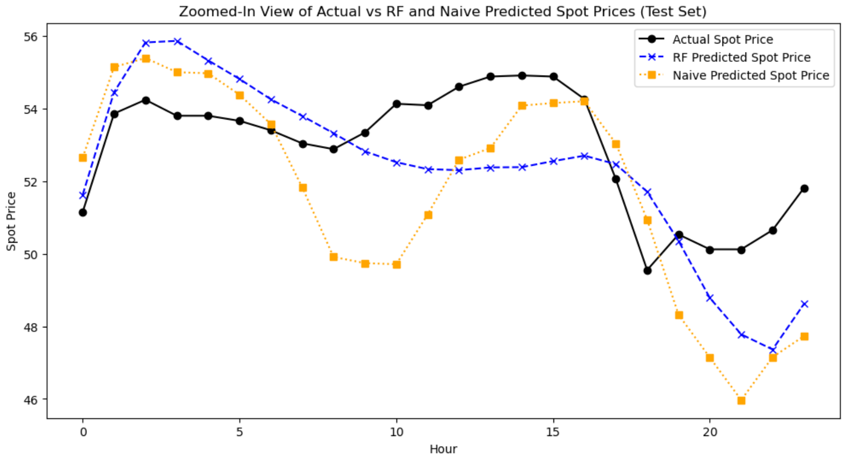


Figure 14: Zoomed-In View of Actual vs. RF and Naive Predicted Spot Prices (Test Set)

4.4.2 RMSE and Error Variability Analysis

To quantify the performance difference, we computed the Root Mean Squared Error (RMSE) for both models. Figure 15 displays the RMSE values, with the RF model achieving a lower RMSE than the baseline, indicating more accurate predictions overall.

Moreover, we examined the standard deviation of errors for each model to assess consistency. As shown in Figure 15, the RF model also demonstrates a lower standard deviation in its errors compared to the baseline, suggesting more stable and reliable predictions over time.

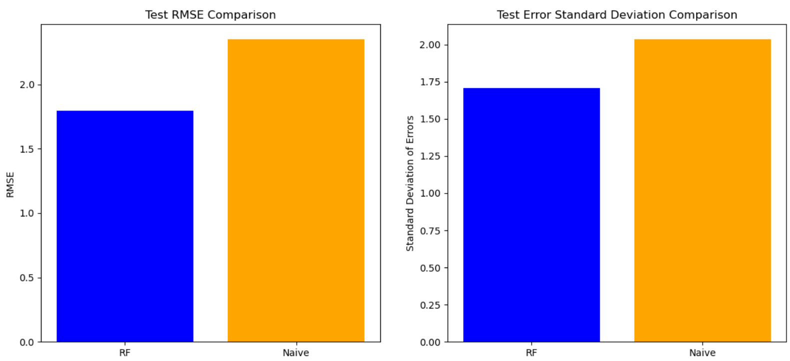


Figure 15: Test RMSE and Standard Deviation of Errors Comparison for RF and Baseline Models

4.4.3 Implications and Observations

The results indicate that the Random Forest model outperforms the baseline in both prediction accuracy and error consistency. While the baseline model provides a simple, intuitive approach by extrapolating from recent values, it lacks the ability to capture more complex relationships in the data. In contrast, the RF model is able to account for non-linear relationships and interactions, making it more suitable for short-term forecasting in the Nord Pool electricity market.

This benchmarking reinforces the utility of the Random Forest model for this forecasting task and justifies its selection over simpler historical baselines. The lower RMSE and reduced error variability translate to more dependable forecasts, which are essential for effective decision-making in energy management.

4.5 Business Value

This section evaluates how our predictive model addresses the defined business objectives for energy producers operating within the Nord Pool market. By meeting these objectives, the model provides actionable insights that drive operational efficiency, strategic market positioning, and financial gains.

4.5.1 Align Production with Demand Trends

Our model's ability to accurately forecast day-ahead spot prices allows energy producers to align their production schedules with expected demand trends. The predictive insight into price fluctuations enables producers to plan production to capitalize on high-demand periods, thereby maximizing revenue. Additionally, improved alignment with demand reduces wasteful overproduction during low-demand periods, contributing to cost efficiency and sustainability.

Future Improvements: To further refine this objective, the model could incorporate more detailed demand predictors, such as real-time weather forecasts or industry-specific demand signals, which would improve responsiveness to sudden demand changes and enhance revenue optimization.

4.5.2 Develop Strategic Bidding Tactics in Day-Ahead Markets

Spot price forecasts support energy producers in formulating competitive and profitable bidding strategies in the Nord Pool day-ahead market. By providing accurate next-day price predictions, the model empowers producers to enter the market at optimal times, maximizing profitability and improving their market positioning. This strategic advantage enables producers to secure better market shares and respond swiftly to favorable market conditions.

Future Improvements: Enhanced model tuning or the integration of additional economic indicators, such as fuel prices or cross-border trading activity, could further refine bidding strategies, allowing producers to adapt bids dynamically and maximize gains in volatile markets.

4.5.3 Enhance Short-Term Planning and Mitigate Risks

With precise spot price forecasts, energy producers can proactively adjust production schedules to avoid unprofitable periods and mitigate exposure to price dips. By identifying potential risks early, producers gain the flexibility to reduce operations during periods of low profitability, thus stabilizing their financial performance and ensuring operational resilience.

Future Improvements: Incorporating a risk assessment module that alerts producers to potential price volatility spikes could further enhance risk mitigation. This module could leverage price sensitivity analysis to identify high-risk periods, providing a more robust decision-making tool for short-term planning.

4.5.4 Optimize Resource Allocation for Operational Efficiency

The model's predictive insights into electricity prices enable producers to allocate resources more effectively by minimizing overproduction and underproduction costs. This optimization not only reduces unnecessary expenditures but also contributes to the sustainability goals of energy producers by preventing resource wastage.

Future Improvements: Adding operational constraints, such as equipment maintenance schedules or resource availability, could help the model optimize resource allocation even further. This would allow producers to maintain production efficiency while balancing operational constraints with market demands.

4.5.5 Increase Profitability from Flexible Assets

For producers with flexible assets like hydro or gas plants, the model's forecast trends enable real-time adjustments in production, allowing them to respond to profitable price spikes. This adaptability ensures that producers can quickly capitalize on favorable market conditions, driving higher profitability for assets with variable output capabilities.

Future Improvements: A dynamic feedback loop could be integrated into the model, enabling it to continuously adjust predictions based on real-time market data. This would provide producers with even more timely insights, enhancing their ability to make profit-maximizing adjustments in real-time.

4.5.6 Summary of Business Value

In summary, our model directly supports Nord Pool energy producers by delivering accurate and actionable spot price forecasts. These forecasts facilitate better production alignment, strategic bidding, risk mitigation, efficient resource allocation, and enhanced profitability from flexible assets. By providing these insights, the model enables energy producers to operate more efficiently, maximize revenue, and adapt swiftly to market fluctuations, reinforcing their competitive positioning in the Nord Pool market.

Each business objective addressed by the model demonstrates its potential for substantial business impact. Future enhancements, such as incorporating additional demand predictors, risk analysis modules, and real-time feedback, would further strengthen the model's effectiveness in supporting strategic decision-making and driving sustained profitability.

4.6 Ensuring the Reliability of the Pipeline

To ensure the effectiveness and reliability of our electricity price prediction model, it is crucial that the entire pipeline, from data processing to prediction generation, is robust and dependable. This pipeline includes data collection, preprocessing, model training, and ultimately, the generation of predictions. Below, we describe the key steps taken to ensure the reliability of our model, which is based on the Random Forest algorithm.

1. **Data Integrity Checks and Preprocessing:** Ensuring data quality is foundational for reliable predictions. Before feeding the data into the model, we conducted comprehensive data integrity checks to identify and handle any missing values, anomalies, or outliers. Automated validation scripts were integrated into the preprocessing pipeline to flag inconsistencies and reduce the risk of feeding inaccurate data into the model.

2. **Use of the Random Forest Algorithm:** The prediction model is built using Random Forest, a machine learning algorithm based on ensembles of decision trees. Random Forest is well-known for its robustness against overfitting and ability to handle large datasets, making

it an ideal choice for predicting electricity prices based on various factors such as demand and production across different hours of the day. The algorithm works by combining the results of multiple decision trees, which reduces the variability of predictions and enhances the overall reliability of the model.

3. **Hyperparameter Tuning and Validation:** To optimize model accuracy, we used Optuna for hyperparameter tuning, and validation splits were applied to assess performance across varied conditions. These methods ensured that the model was not overly tailored to historical data, enhancing its generalizability and preventing overfitting.

4. **Pipeline Scalability and Automation:** To ensure scalability, our pipeline is designed to handle increased data volumes as the electricity market grows or more variables are incorporated. Automation steps were introduced, including automatic retraining and model updating, allowing the pipeline to adapt seamlessly to new data without requiring manual intervention. This automation is crucial for maintaining model accuracy over time, especially as market conditions evolve.

5. **Performance Monitoring and Drift Detection:** We implemented performance monitoring to continually evaluate model accuracy and identify any signs of data or model drift. By tracking changes in error metrics over time, we can detect when the model's predictions start deviating significantly from actual prices, indicating a need for retraining or adjustment. This proactive approach helps maintain the model's reliability even as market conditions shift.

6. **Scalability and Future Adaptability:** The pipeline is built with adaptability in mind, allowing for easy integration of additional data sources or expanded prediction requirements, such as multiple time steps. This flexibility ensures that the model can meet evolving business needs without extensive reconfiguration.

This approach emphasizes the importance of data quality, model robustness, and proactive monitoring to ensure a reliable and scalable forecasting system. With these steps, the pipeline is designed to be dependable in real-world conditions and adaptable to future needs.

5 Deployment and Recommendations

5.1 Deployment Plan

To deploy our electricity price prediction model, originally designed for SINTEF, we will use an API (Application Programming Interface). This API will allow SINTEF and other potential clients to access our predictions efficiently, scalably, and securely, facilitating the integration of our data into their systems quickly and directly. This deployment plan details each necessary step to design, develop, and implement this API, covering aspects such as endpoint configuration, security, scalability, and continuous monitoring. By providing access to quality predictions through an API infrastructure, we ensure efficient delivery of our data, laying the foundation for a satisfied and continuously growing client network. Definition of Objectives and Requirements

1. **Main Objective:** Provide our clients with electricity price predictions through an easily accessible, secure, and scalable API.

2. **Technical requirements:**

   We must ensure the API supports a high volume of concurrent requests without affecting system performance. We have to implement authentication and authorization mechanisms to protect prediction data. Besides of that, we should define performance metrics to evaluate the success of the API (response time, uptime, etc.).

3. **API architecture design platform and framework selection:** The framework for the API development could be Flask or FastAPI (for Python), because the code has been done with this language and it is ideal for fast and efficient APIs. Django REST Framework (for

larger-scale projects in Python). In a cloud hosting environment a cloud provider such as AWS or Google Cloud could be useful. For the API on a cloud service, we should ensure a proper server configuration to handle expected traffic.

4. **Authentication and security user authentication:** Implement token-based authentication, such as OAuth2 or API Keys, to ensure that only authorized users can access the API. We are going to use HTTPS to encrypt data in transit and protect sensitive information.

5. **Request limits and access control:** We should configure a rate limiting to prevent abuse and a good working of the system, furthermore, all the log in requests should be monitored on a system to track and prevent malicious access.

6. **Load and performance testing:**

   Conduct load testing to verify that the API can handle the volume of requests without performance degradation. Adjust the infrastructure based on test results to ensure continuous availability. Automatic Scalability: Configure automatic scalability on the server to increase capacity in case of a sudden increase in demand (e.g., AWS Auto Scaling) in a growth of clients scenario.

7. **Performance monitoring:**

   The use of monitoring tools (such as AWS CloudWatch or New Relic) to observe key metrics (response time, error rates, memory and CPU usage) is needed. Then, setting up alerts to notify about performance issues or service outages would help to avoid system failures .

8. **Model and data updates:**

   Periodically update the prediction model with recent data to ensure predictions are accurate and reflect current market conditions. In order to avoid a performance degradation a egression testing before each update should be done.

9. **Cost optimization:**

   Review and adjust cloud infrastructure to optimize costs without sacrificing performance, based on actual API usage.

10. **Collecting client feedback:** Request feedback from clients to identify possible improvements in the API, such as adding new endpoints or improving response times.We should review usage patterns to better understand how clients interact with the API and adjust resources or functionalities according to their needs.

5.1.1 Benefits of Implementing the API

1. **Real-Time access:** Clients receive updated prediction data in real-time, allowing them to make informed decisions quickly.

2. **Efficiency and scalability:** The API enables multiple clients to simultaneously access predictions, optimizing response time and facilitating integration with other systems.

3. **Security and control:** With authentication and encryption, data is protected, providing clients with peace of mind regarding secure access to information.

4. **Process automation:** Clients can integrate the API into their own systems, automating price management and risk reduction tasks.

This plan allows the company to offer electricity price predictions efficiently, securely, and scalably, aligning with client needs and optimizing its own workflow to manage the service continuously and effectively.

5.2 Relevant Actions Towards Different Stakeholders

This section will detail relevant actions different stakeholders are recommended to take based on results from our analysis, as well as suggestible actions based on the model predictions. Since we conducted this project for SINTEF, but the model and predictions are made for one of SINTEF clients (Nord pool),we can differentiate between SINTEF as a stakeholder and the client as a stakeholder. There are actions that can be taken by SINTEF to make their consumption prediction service as attractive as possible to their clients, and there are actions the clients can take based on the provided predictions.

5.2.1 SINTEF Action

The company that owns or manages this price prediction model can take various strategic and operational measures based on the model to maximize its profits and improve resource management.

1. **Regular Model Validation:** To ensure the model maintains good performance over time, SINTEF should implement frequent validation and backtesting procedures. This will help identify potential deviations in model accuracy and allow for adjustments to maintain its reliability and usefulness in changing market scenarios.
2. **Developing interpretability tools for decision making:** Given the complexity of a price prediction model, SINTEF could provide its clients with interpretability tools, such as feature importance visualizations or prediction explanations. This would allow their clients to better understand the factors influencing predictions and make more informed decisions based on the model.
3. **Training and support for SINTEF's clients :** To maximize the model's value, SINTEF should offer training and technical support to Nord Pool and other clients. This will help users understand the model, its limitations, and the best ways to interpret it, thereby increasing customer satisfaction and confidence in the predictions.
4. **Collecting feedback from clients and end Users:** Feedback from end users is crucial for identifying areas of improvement. SINTEF could establish a system to collect client feedback on the model's performance and accuracy, allowing for continuous improvements and adaptations based on specific user needs.

5.2.2 Client Actions

The main client, Nord pool, could benefit in different ways from our electricity price prediction model, in addition to indirectly benefiting other companies or the electricity market itself,by contributing to a more stable and efficient electricity market.

1. **Market optimization:** With more accurate predictions, Nord Pool can adjust its operations to more accurately reflect real-time market conditions. This allows Nord Pool to improve pricing efficiency, benefiting all market participants. A price that better reflects actual supply and demand creates a more transparent and reliable market, increasing user confidence and fostering a fairer and more effective trading environment.
2. **Risk reduction:** By more accurately anticipating price fluctuations, Nord Pool can better manage the risks associated with market volatility. This is crucial in a dynamic market like electricity, where price changes can have a significant financial impact on participants. With a robust prediction model, Nord Pool can implement strategies to mitigate these risks, protecting both its own stability and that of its clients.
3. **Improved planning:**More accurate price predictions facilitate the planning of supply and demand in the electricity market. Nord Pool can ensure there is sufficient capacity to meet market needs without incurring additional costs or an inefficient surplus. This optimization

of planning reduces the risks of shortages or excess capacity, improving market stability and minimizing resource waste.

4. **Promoting competition:**By providing more accurate and timely information, Nord Pool can help create a fairer and more efficient competitive environment among market participants. When all actors have access to reliable price predictions, they can make informed decisions and compete on equal terms. This not only fosters healthier competition but also helps maintain fair prices for end consumers.

5. **Integration of renewable energy:**An effective prediction model also allows Nord Pool to better integrate renewable energy sources. Since renewable energy production (such as solar or wind) is more variable and depends on external factors like weather, an accurate prediction model helps anticipate these variations and adjust supply accordingly. This facilitates the integration of renewable energies into the market, promoting a transition to a more sustainable energy system and helping to meet European sustainability goals.

   Overall, these benefits enable Nord Pool to optimize its operations, reduce financial risks, improve operational efficiency, and contribute to a more stable, sustainable, and competitive electricity market. This not only strengthens Nord Pool's position in the market but also provides added value for all its participants and supports a transition to cleaner and more accessible energy for everyone.

5.3 Limitations and Improvements

5.3.1 Dataset Limitations and Improvements

To improve the database, more aspects of those already provided would have to be taken into account so that the prediction model adjusts even more to reality and then acts more effectively and reliably.

1. **Incorporating additional and exogenous data:** SINTEF could consider adding weather conditions, renewable energy availability and fuel prices. Adding socioeconomic data that affect electricity demand, such as economic growth or industrialization indicators in the region could be a great addition too. These factors would provide a more comprehensive perspective on the variability in electricity demand and, therefore, improve prediction accuracy. Adding political decisions, value of raw materials or available nuclear energy,would also help to improve the prediction model.

2. **Frequency and updating of data:** It is essential that the dataset is continuously updated. Electricity prices are highly sensitive to real-time changes, such as fluctuations in renewable energy production or unexpected consumption. SINTEF could benefit from regularly updating its database with the most recent and high-frequency data sources, so the model more accurately reflects current market conditions.

3. **Quality and accuracy of training data:** Data quality is a critical factor for model accuracy. SINTEF must ensure that the data is free from errors, inconsistencies, or outliers that could affect predictions. Implementing data cleaning and quality verification techniques can reduce noise and improve the model's robustness.

4. **Possible bias in historical data:** Historical price data may be influenced by specific events or temporary policies that will not recur in the future. SINTEF should consider techniques to detect and correct potential biases in the training data, so the model is not disproportionately influenced by anomalous events that do not represent real patterns.

5.3.2 Method Limitations and Improvements

Developing our electricity price prediction model, we achieved notable accuracy and consistency in our results. However, as with any predictive modeling approach, there are certain limitations that must be acknowledged to provide a balanced perspective on the model's performance and potential applications. Recognizing these limitations is essential for understanding the model's current constraints, areas where it might be improved, and the contexts in which its predictions may be less reliable.

In this section, we discuss key limitations related to the model's structure, data features, interpretability, and generalization capacity. Identifying these issues not only clarifies the model's current range of applicability but also highlights directions for future refinement and potential adaptation to broader contexts in energy forecasting.

1. **Lack of Additional Lag Features:** During data analysis, we observed that autocorrelation in our dataset significantly decreases after approximately 24 hours, suggesting that past values have less influence on predictions beyond this timeframe. Since we have data covering only the past two years, adding additional lag features might not be fully relevant. However, an alternative approach could involve generating synthetic data for earlier time periods, which could strengthen the model by better utilizing correlations within the significant 24-hour window.

2. **Lack of Confidence Estimates in Predictions:** Our current model does not automatically provide confidence intervals or uncertainty estimates in its predictions. This limitation could be significant in a context where prediction errors can result in high costs for energy management. Incorporating methods to calculate uncertainty in each prediction would help improve decision-making accuracy, giving energy resource managers a more comprehensive view of the reliability of forecasts.

3. **Improvement in Feature Selection:** While the current model incorporates factors such as hourly production and demand, it may not fully capture all relevant variables affecting electricity prices. Additional variables, such as extreme weather conditions or fluctuations in socioeconomic factors, could provide valuable insights into pricing patterns. Expanding the set of features could provide a more accurate and holistic view of the energy system.

4. **Model Complexity:** Using a single predictive model limits the diversity of patterns captured from the data and increases the risk of overfitting. Ensemble approaches, which combine multiple models, often improve accuracy and adaptability by balancing individual errors and variability. Without the advantages of an ensemble approach, the model may be less adaptable to changes in data over time, potentially resulting in reduced accuracy and robustness across different conditions.

5. **Limited Generalization:** The model is optimized for the specific conditions of the dataset used, which could limit its applicability in other contexts or energy markets with different characteristics. This limitation on generalization means the model would require adjustments or retraining to adapt to new scenarios, increasing costs and reducing effectiveness in diverse situations.

5.4 Future Analysis

Some key points to guide our future analysis and complement the results of the electricity price prediction model have to be taken into account.

1. **Incorporation of alternative models and performance comparison:** Testing different machine learning algorithms, such as Gradient Boosting (XGBoost, LightGBM) or neural network models, could improve the model's accuracy and robustness. Conducting a comparative analysis between these models and the current Random Forest model would help identify the best approach for the price prediction problem.

2. **Evaluation of model interpretability:** Investing in interpretability techniques such as SHAP (Shapley Additive Explanations) or LIME (Local Interpretable Model-agnostic Explanations) can help better understand the model's decisions and identify which factors most

influence the predictions. This is particularly valuable for helping clients understand how and why the model makes certain predictions.

3. **Continuous model monitoring and updating:** Establishing a monitoring system to evaluate the model's performance in production and detect possible deviations in prediction accuracy over time. This includes updating the model with new data and periodically adjusting hyperparameters to ensure the model remains relevant and accurate in a changing market.

4. **Scenario and sensitivity analysis:** Creating analysis scenarios to evaluate how the model responds to extreme events, such as abrupt changes in demand or supply disruptions. This will allow anticipation of how the model might respond in unexpected situations and adapt decision-making accordingly. Additionally, a sensitivity analysis will help understand the impact of each variable on the final prediction.

5. **Integrating user feedback:** Consistently gathering and incorporating feedback from the end-users of the forecasting system can offer valuable insights for refining the model to better align with the practical needs of the business.

6 Monitoring and Maintenance

The effectiveness of a predictive model for electricity spot price forecasting relies not only on its initial accuracy but also on its ability to remain reliable and relevant over time. Given the dynamic nature of the Nord Pool electricity market, continuous monitoring and maintenance are essential to ensure that the model adapts to evolving trends, seasonality, and any unexpected fluctuations in demand or production [AHM20]. Research indicates that models in production require regular monitoring to detect issues like data drift, which can degrade performance if not addressed through retraining or parameter adjustments [Gam+14].

Our approach to monitoring and maintenance will encompass both technical metrics and business-relevant KPIs, ensuring the model remains aligned with the needs of stakeholders. Automated alerts, dashboards, and feedback loops will facilitate prompt action in response to any performance declines, while regular model retraining and data quality checks will uphold model accuracy and relevance [Com23]. This proactive approach aims to maximize the model's value as a decision-support tool for energy producers, traders, and other market participants in the Nord Pool environment.

6.1 Key Performance Indicators

To maintain the model's effectiveness in forecasting electricity spot prices, we rely on key performance indicators (KPIs) that gauge both technical accuracy and business impact. These KPIs provide stakeholders with insights into the model's reliability and guide necessary adjustments for sustained performance.

Our KPIs are structured into two categories: Model Performance KPIs, which focus on technical accuracy, and Business KPIs, which evaluate the model's impact on operational efficiency and strategic decision-making. By tracking these indicators, we proactively align the model with market dynamics and stakeholder needs, ensuring it remains a valuable asset.

6.1.1 Model Performance KPIs

Technical performance is critical for an effective forecasting tool, especially in the Nord Pool market, where accuracy directly affects financial outcomes. The following KPIs help us monitor and maintain the model's reliability:

1. **Mean Squared Error (MSE)**: As the primary accuracy metric, MSE measures the average squared difference between predicted and actual prices, emphasizing larger errors. This focus on substantial deviations is crucial for maintaining accuracy in price forecasting, where even minor errors can impact scheduling and bidding strategies.
2. **Baseline Comparison**: To validate the added value of the Random Forest model, we compare its performance against a naive baseline that replicates the previous day's prices. This comparison highlights the model's improvement over simpler methods, ensuring stakeholders see clear value in using a more complex model.
3. **Rolling Mean Squared Error**: A rolling MSE metric captures performance trends over recent periods (e.g., weekly). This approach helps detect seasonal fluctuations, data drift, or market shifts that may require retraining or recalibration, supporting proactive model maintenance.

These metrics provide a comprehensive view of technical accuracy, facilitating ongoing assessment for both technical teams and stakeholders.

6.1.2 Business KPIs

In addition to technical performance, Business KPIs assess the model's alignment with operational and strategic goals. These KPIs reveal the model's impact on decision-making and risk management within the Nord Pool market:

- **Forecast Accuracy for High-Demand Periods**: Accurate predictions during peak demand help optimize production schedules and prevent costly imbalances, enhancing efficiency and reducing financial risk.
- **Market Bidding Strategy Effectiveness**: This KPI tracks the model's support in shaping effective bidding strategies, aligning predictions with actual price trends to improve profitability and reduce exposure to unfavorable price movements.
- **Price Volatility Sensitivity**: Monitoring accuracy during volatile periods helps stakeholders gauge the model's robustness under fluctuating market conditions, ensuring reliable support even during unpredictable events.

By using these KPIs, stakeholders can continually assess the model's operational and strategic contributions. We also incorporate feedback from stakeholders to refine these KPIs over time, adapting to changes in the market and business objectives.

6.2 Monitoring Dashboard

A well-constructed monitoring dashboard is essential for our operational team to maintain a real-time pulse on the forecasting model's performance. This dashboard provides an intuitive and customizable interface, allowing quick assessment of critical Key Performance Indicators (KPIs) and offering predictive insights to support operational decision-making.

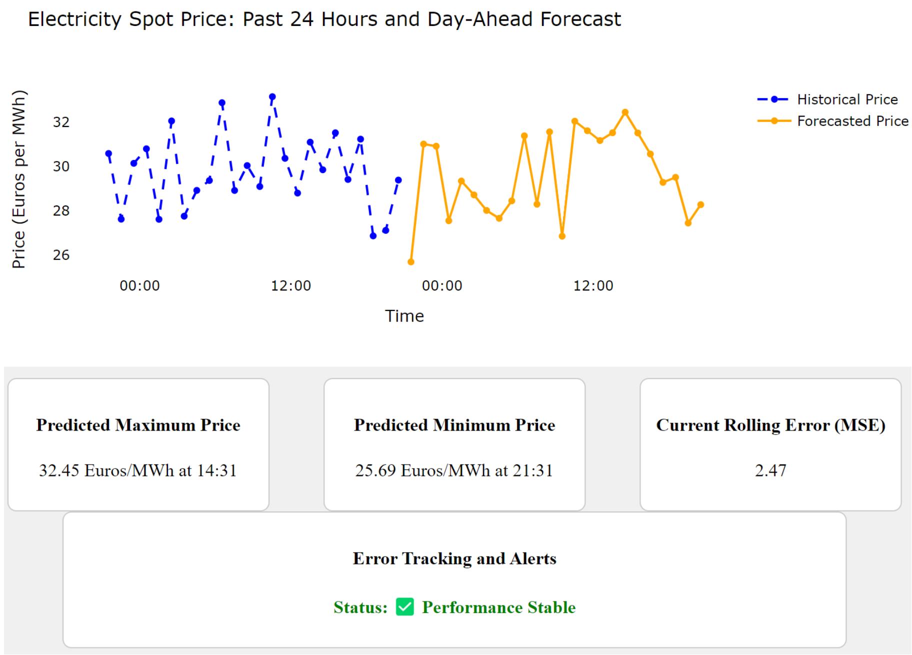


Figure 16: Our Electricity Spot Price Forecasting Dashboard

Figure 16 illustrates our dashboard, which includes the following components:

- **Main Forecast Visualization**: The central line chart displays the day-ahead electricity spot price predictions along with historical prices from the previous 24 hours. This visualization allows stakeholders to observe price trends and identify potential anomalies in real time, supporting Business KPIs such as Forecast Accuracy for High-Demand Periods and Price Volatility Sensitivity.
- **Performance Summary (Key Metrics)**: Below the forecast visualization, key metrics such as the predicted maximum and minimum spot prices, their respective times, and the current rolling Mean Squared Error (MSE) are displayed. The rolling MSE metric provides a snapshot of recent forecast accuracy, helping the team detect performance changes due to evolving market conditions.
- **Error Tracking and Alerts**: The dashboard includes an alert status to notify stakeholders if the model's MSE or baseline comparison exceeds a predefined threshold. This proactive alert feature enables rapid response to potential issues, ensuring model performance remains aligned with both technical and business KPIs.

The dashboard serves as a proactive tool, enabling operational teams to monitor, manage, and anticipate system performance and potential forecasting issues. By offering real-time insights and customizable KPIs, it helps maintain forecast accuracy and supports business scalability.

6.3 Strategic Risk Management and Contingency Planning

Deploying a machine learning model in the context of electricity price forecasting requires careful risk management and proactive planning to maintain stable performance over time. This section outlines the key risks and corresponding contingency strategies to ensure the robustness of our forecasting model in a dynamic market environment.

- **Model Performance Degradation**: As electricity pricing patterns may change due to seasonal shifts, regulatory updates, or unexpected economic factors, there is a risk that the model's accuracy could degrade over time. To mitigate this, regular monitoring of model performance KPIs, such as Mean Squared Error (MSE) and rolling MSE, will be essential. Persistent deviations in prediction accuracy should prompt either a retraining of the model with more recent data or a review of the model's architecture to incorporate any new influential factors, ensuring the model remains aligned with current market conditions.

- **Infrastructure Failures**: Our forecasting model relies on cloud-based infrastructure, which brings risks of service downtime or other technical disruptions. To manage this, we recommend establishing a backup plan with an alternative server or cloud provider that can be activated if our primary service experiences outages. Furthermore, local deployment of a minimal model version could be utilized as a temporary solution during brief service interruptions, minimizing any potential disruption to stakeholders.
- **Data Security and Privacy Risks**: Given the need to handle sensitive market and price data, the risk of unauthorized access or data breaches must be addressed. Our strategy includes implementing security protocols such as data encryption, access controls, and regular security audits to detect vulnerabilities. Additionally, we recommend setting up alerts for unusual access patterns, enabling the team to act swiftly in the event of suspicious activity or potential security breaches.
- **Market and Regulatory Changes**: The electricity market is heavily regulated, and changes in regulatory policies may affect data handling requirements or model specifications. To ensure compliance, our team will closely monitor any relevant regulatory developments and adjust the model or data processing procedures as necessary. Additionally, maintaining a flexible model architecture will allow us to adapt quickly to any new compliance requirements.
- **Human Oversight and Intervention**: Despite the model's automation capabilities, human oversight remains essential to respond to unexpected scenarios and manage any alerts generated by the monitoring system. A dedicated operations team will be responsible for overseeing model performance, investigating alerts, and performing quick troubleshooting when needed. Regular communication with stakeholders will ensure transparency about significant alerts and interventions. To support this, thorough documentation and ongoing training will be provided, empowering the team to effectively manage issues and maintain optimal performance.

Incorporating these risk management and contingency strategies will help ensure that our forecasting model remains reliable, secure, and adaptable in a complex and evolving market environment. Through proactive monitoring, stakeholder feedback, and response protocols, the model will be well-positioned to support accurate forecasting and facilitate informed decision-making.

6.4 Lessons Learned and Feedback Integration

This project underscored the importance of integrating technical expertise with a comprehensive understanding of client needs and market dynamics. Developing a machine learning model for electricity price forecasting required not only technical proficiency but also an acute focus on aligning model capabilities with stakeholder requirements and business objectives. Bridging this gap between advanced technology and practical usability proved essential to the project's success.

Adopting a user-centric approach helped us design a model that was both high-performing and actionable for stakeholders. This focus on stakeholder needs guided our decisions on model design, monitoring strategies, and interpretability, aligning with Design Thinking principles that emphasize iterative refinement and adaptability to real-world challenges. By prioritizing user accessibility, we ensured that the model not only met performance standards but also supported decision-making effectively in the dynamic Nord Pool market.

The Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology was instrumental in structuring our work, allowing us to approach each project phase systematically. From initial problem understanding to solution deployment, CRISP-DM helped us maintain a balanced focus on technical and business aspects, ensuring that each phase aligned with the goal of creating a robust, valuable model. For instance, the "Data Understanding" phase revealed key insights into the importance of high-demand periods and price volatility, which influenced our focus on business KPIs, while "Model Evaluation" emphasized the importance of ongoing monitoring through performance metrics.

The continuous feedback loop established throughout the project was particularly valuable, as it allowed us to adapt the model based on real-time performance metrics and stakeholder input. The dual focus on both technical KPIs (like MSE) and business KPIs (such as Market Bidding Strategy Effectiveness) enabled us to evaluate the model's success not only by its accuracy but also by its strategic impact on operations. This structure proved effective for anticipating market changes and addressing model performance issues proactively, ensuring that the model continued to meet evolving business needs.

Moving forward, this feedback-informed, KPI-driven approach will serve as a cornerstone for continuous improvement, helping to maintain the model's relevance and effectiveness in a complex and shifting market environment. Overall, this project highlighted the need for a disciplined yet flexible approach to machine learning. By aligning technical development with stakeholder priorities and maintaining room for iterative improvements, we achieved a model that provides accurate, actionable forecasts while remaining responsive to the demands of an evolving market.

Bibliography

[AHM20] Rehan Ahmed, Farookh Hussain and Saad Mostafa. 'Machine Learning in Energy Systems: A Review'. In: *IEEE Access* 8 (2020), pp. 164644–164665.

[arX24] arXiv. *Energy Price Modelling: A Comparative Evaluation of Four Generations of Models*. Accessed: 2024-11-14. Nov. 2024. URL: https://arxiv.org/abs/2411.03372.

[Aus23] Australian Energy Market Operator (AEMO). *Electricity Forecasting Insights*. Accessed: 2024-11-14. 2023. URL: https://aemo.com.au/Electricity/Electricity-forecasting-insights.

[Bre01] Leo Breiman. 'Random Forests'. In: *Machine Learning* 45.1 (2001), pp. 5–32.

[Bro23a] Jason Brownlee. *How to Forecast Multiple Steps Ahead in Time Series Forecasting*. Accessed: 2024-11-14. 2023. URL: https://machinelearningmastery.com/multi-step-time-series-forecasting/.

[Bro23b] Jason Brownlee. *Random Forest for Time Series Forecasting*. Accessed: 2024-11-14. 2023. URL: https://machinelearningmastery.com/random-forest-for-time-series-forecasting/.

[BU16] Walter Brenner and Falk Uebernickel, eds. *Design Thinking for Innovation: Research and Practice*. Library of Congress Control Number: 2016933122. Cham, Switzerland: Springer, 2016. ISBN: 978-3-319-26098-3. DOI: 10.1007/978-3-319-26100-3. URL: https://doi.org/10.1007/978-3-319-26100-3.

[Cle+90] Robert B. Cleveland et al. 'STL: A Seasonal-Trend Decomposition Procedure Based on LOESS'. In: *Journal of Official Statistics* 6.1 (1990), pp. 3–73.

[Com23] Machine Learning Operations Community. *Monitoring and Maintaining Machine Learning Models in Production*. Accessed: 2024-11-14. 2023. URL: https://mlops.community/resources/monitoring-maintaining-ml-models/.

[Gam+14] Joao Gama et al. 'A Survey on Concept Drift Adaptation'. In: *ACM Computing Surveys* 46.4 (2014), pp. 1–37.

[HTF09] Trevor Hastie, Robert Tibshirani and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd. Springer, 2009.

[Int22] International Energy Agency. *World Energy Outlook 2022*. Accessed: 2024-11-14. Nov. 2022. URL: https://www.iea.org/reports/world-energy-outlook-2022.

[Int23] International Energy Agency. *Digitalization and Energy*. Accessed: 2024-11-14. June 2023. URL: https://www.iea.org/reports/digitalization-and-energy.

[MDP23] MDPI. 'Data-Driven Techniques for Short-Term Electricity Price Forecasting'. In: *Energies* 17.7 (2023). Accessed: 2024-11-14, pp. 1625–1640. URL: https://www.mdpi.com/1996-1073/17/7/1625.

[Nor24a] Nord Pool Group. *About Nord Pool - Our Business*. Accessed: 2024-11-14. Nov. 2024. URL: https://www.nordpoolgroup.com/en/About-us/our-business/.

[Nor24b] Nord Pool Group. *Day-ahead Market*. Accessed: 2024-11-14. Nov. 2024. URL: https://www.nordpoolgroup.com/en/the-power-market/Day-ahead-market/.

[Nor24c] Nord Pool Group. *The Power Market*. Accessed: 2024-11-14. Nov. 2024. URL: https://www.nordpoolgroup.com/en/the-power-market/.

[She00] C. Shearer. 'The CRISP-DM model: the new blueprint for data mining'. In: *Journal of Data Warehousing* 5.4 (2000), pp. 13–22.

[TAH20] Youssef Tounsi, Houda Anoun and Larbi Hassouni. *CSMAS: Improving Multi-Agent Credit Scoring System by Integrating Big Data and the new generation of Gradient Boosting Algorithms*. Mar. 2020. DOI: 10.1145/3386723.3387851.

[Wer14] Rafał Weron. *Electricity price forecasting: A review of the state-of-the-art with a look into the future*. Accessed: 2024-11-14. Apr. 2014. URL: https://doi.org/10.1016/j.ijforecast.2013.09.005.

[Wor23] Renewable Energy World. *Renewable Energy Forecasting: Why It Matters*. Accessed: 2024-11-14. May 2023. URL: https://www.renewableenergyworld.com/articles/2023/renewable-energy-forecasting-why-it-matters/.

[ZPZ21] Jie Zhang, Yanlin Peng and Lei Zhang. 'Recursive Strategies in Multi-Step Time Series Forecasting'. In: *IEEE Access* 9 (2021), pp. 12345–12356. URL: https://ieeexplore.ieee.org/document/10331086.

Use of AI:

Chatbots like ChatGPT and Gemini have been used as a support to program the model and help us creating the graphics and figures. Moreover, we have used chatGPT to re-phrase some parts of our writing. This was to bridge the gap in how each individual writes, so the full document felt more cohesive. *Github Copilot* has been also used for this task.