Towkir Ahmed

# Enhancing Price Forecasting Accuracy in Reserve Markets with a focus on addressing the challenges in forecasting FCR-N prices for BSPs using data from Fingrid's open data platform

School of Technology and Innovations

Master's Thesis in Smart Grids Programme

Vaasa 2024

**VAASAN YLIOPISTO**
**School of Technology and Innovations**
**Author:** Towkir Ahmed
**Thesis title:** Enhancing Price Forecasting Accuracy in Reserve Markets with a focus on addressing the challenges in forecasting FCR-N prices for BSPs using data from Fingrid's open data platform
**Degree:** Master of Science in Technology
**Discipline:** Smart Grids
**Supervisor:** Hannu Laaksonen, Hosna Khajeh
**Year of graduation:** 2024 **Number of pages:** 47

**ABSTRACT:**
This thesis investigates machine learning techniques to improve forecasting accuracy for Frequency Containment Reserve for Normal operation (FCR-N) prices in reserve markets, with a focus on Finland's market data obtained from Fingrid's open data platform. Given the importance of accurate price predictions for Balancing Service Providers (BSPs), the study evaluates multiple forecasting models, including regression-based, gradient boosting, and neural network approaches, to address the complexities in reserve pricing. Key features, such as non-solar and non-wind generation and FCR-N price lags, were incorporated as predictors, while solar and wind forecasts were excluded due to their low correlation with FCR-N prices. Results indicate that models like LightGBM and recurrent neural networks (RNNs) demonstrate high predictive accuracy, capturing temporal trends effectively. However, limitations, including data granularity and model interpretability, suggest areas for future work, such as integrating external market factors and improving real-time adaptability. This study underscores the potential of machine learning to enhance price forecasting in reserve markets, aiding BSPs in strategic decision-making and supporting more robust energy management in the evolving energy landscape.

**Keywords:** FCR-N Pricing, Time Series Forecasting, Energy Economics

# Contents

List of Figures 5

1 Introduction 7

1.1 Background 7

1.2 Research Questions 9

1.3 Research delimitations 9

1.4 Research Plan and Structure 10

2 Literature Review 12

2.1 Reserve Market Price Forecasting 12

2.2 Machine Learning Approaches in Electricity Price Forecasting 14

2.3 Data-Driven Approaches and Open Data Utilization 15

2.4 Nordic Market Specific Studies 16

2.5 Gap in Current Research 18

3 Dataset 19

3.1 Data Collection 19

3.2 Use of Forecasted Values for Consumption, Solar, and Wind 20

3.3 Data Visualization 21

4 Methodology 24

4.1 Data Preprocessing 24

4.1.1 Resampling 24

4.1.2 Merging the datasets 24

4.1.3 Defining non-solar and non-wind generation 24

4.1.4 Calculating correlations 25

4.1.5 Getting the final data 27

4.1.6 Scaling the data 27

4.1.7 Splitting the data for training 28

4.2 Machine Learning Techniques 29

4.2.1 Regression based Models 29
4.2.2 Gradient Boosting based Models 31
4.2.3 Neural Network based Models 32

5 Result Analysis 34
5.1 Model Performance Evaluation: 34
5.2 Comparative Analysis of Forecast Accuracy 36

6 Discussion 40
6.1 Theoretical Implications 40
6.2 Limitations 40
6.3 Future Research Possibilities 41

7 Conclusion 43

Bibliography 45

## List of Figures

Figure 1 Time series plot for FCR-N prices 21
Figure 2 Time series plot for Electricity consumption forecast 22
Figure 3 Time series plot for Wind and Solar power generation forecast 23
Figure 4 Autocorrelation for FCR-N prices for lags 24-48 26
Figure 5 Pearson correlation for non-solar and non-wind generation and FCR-N prices 27
Figure 6 Spearman correlation for non-solar and non-wind generation and FCR-N prices 28
Figure 7 Cross-Correlations (Pearson) 29
Figure 8 Cross-Correlations (Spearman) 30
Figure 9 Performance comparison of all models 35
Figure 10 The training values of FCR-N prices and the RNN predictions of FCR-N prices on the test set results. 36
Figure 11 Comparison of RNN model predictions (orange) with actual FCR-N prices (blue) 36
Figure 12 The training values of FCR-N prices and the LSTM predictions of FCR-N prices on the test set results. 37
Figure 13 Comparison of LSTM model predictions (orange) with actual FCR-N prices (blue) 37
Figure 14 The training values of FCR-N prices and the GRU predictions of FCR-N prices on the test set results. 38
Figure 15 Comparison of GRU model predictions (orange) with actual FCR-N prices (blue) 38

## Abbreviations

| | |
|---|---|
| AI | Artificial Intelligence |
| FCR-N | Frequency Containment Reserve for Normal operation |
| BSP | Balancing Service Providers |
| TSO | Transmission System Operators |

| | |
|---|---|
| RNN | Recurrent Neural Networks |
| GRU | Gated Recurrent Unit |
| LSTM | Long Short Term Memory |

# 1 Introduction

This section provides the overall intuition behind the research. The background introduces the importance of the FCR-N market for grid stability, where BSPs rely on price forecasts to make strategic decisions. This section highlights the challenges of forecasting due to market volatility and how machine learning can enhance prediction accuracy using Fingrid's open data. Research questions outlines key issues: identifying factors influencing FCR-N prices, evaluating machine learning model accuracy, and determining the best forecasting model. These questions guide the study toward understanding FCR-N price dynamics. Research delimitations subsection discusses limitations such as data quality issues, model assumptions, external market influences, and computational constraints, all of which may impact forecasting accuracy. Lastly, research plan and structure outlines the thesis timeline: data collection, feature engineering, model development, evaluation, and writing. Each phase is structured to ensure systematic analysis and timely completion of the research.

## 1.1 Background

In the rapidly evolving landscape of electricity markets, the accurate forecasting of reserve market prices has become increasingly crucial for both system operators and market participants. Reserve markets play a crucial role in maintaining grid stability by ensuring that sufficient energy reserves are available to meet unexpected fluctuations in supply and demand. Among these, the Frequency Containment Reserve for Normal operation (FCR-N) plays a vital role in maintaining grid stability, reliability and serves as an essential product in the Nordic electricity grid. Managed by Transmission System Operators (TSOs) such as Fingrid in Finland, FCR-N helps control the frequency of the power grid and provides a buffer against imbalances. The hourly prices in this market are subject to fluctuations based on supply, demand, and cross-border procurement activities.

As power systems transition towards greater integration of renewable energy sources, the volatility and unpredictability of FCR-N prices pose significant challenges for Balancing Service Providers (BSPs) in optimizing their market strategies and operational decisions. That is why, for Balancing Service Providers (BSPs), accurately forecasting the FCR-N prices is essential, as it allows them to better manage costs, optimize their reserve bids, and maximize profitability. The dynamic nature of reserve markets, particularly influenced by factors like renewable energy production, load demands, and inter-regional energy trades, makes accurate price forecasting challenging. The Finnish transmission system operator, Fingrid, has taken a progressive step by providing an open data platform, offering unprecedented access to historical market data. This initiative presents a unique opportunity to develop and refine price forecasting models for the FCR-N market. With Fingrid's extensive open data platform, researchers and industry players can access comprehensive datasets that include real-time and historical data on energy consumption, renewable energy production, and reserve prices, thus opening opportunities to enhance forecasting methodologies. However, despite the availability of this rich dataset, accurately predicting FCR-N prices remains a complex task due to the multitude of factors influencing reserve markets, including weather patterns, demand fluctuations, and the increasing penetration of intermittent renewable energy sources.

This thesis aims to bridge this gap by developing an enhanced price forecasting model for the FCR-N market, specifically tailored to the needs of BSPs operating in the Finnish electricity market. By leveraging Fingrid's open data platform and employing state-of-the-art machine learning algorithms, this research seeks to improve the accuracy and reliability of FCR-N price predictions. The study will explore various forecasting methodologies, including time series analysis, artificial neural networks, and ensemble methods, to identify the most effective approach for capturing the unique characteristics of FCR-N price dynamics. The outcome will provide valuable insights for BSPs, helping them better understand price drivers and make informed bidding and procurement decisions.

## 1.2 Research Questions

This study addresses several research questions designed to enhance our understanding of FCR-N price dynamics and improve forecasting capabilities for BSPs:

- What are the primary features influencing FCR-N price fluctuations in Fingrid's reserve market?
- How accurately can various machine learning algorithms forecast FCR-N prices using Fingrid's open data?
- Which machine learning model provides the best balance of accuracy and computational efficiency for FCR-N price forecasting?
- How do external factors, such as renewable energy generation and cross-border trades, impact FCR-N price predictability?

These questions aim to dissect the complexities of price forecasting in the FCR-N reserve market and provide actionable insights into feature selection and model performance.

## 1.3 Research delimitations

While this research aims to improve FCR-N price forecasting accuracy, several limitations must be acknowledged:

- **Data Availability and Quality:** Although Fingrid's open data platform provides extensive datasets, occasional missing or erroneous values could impact the model's accuracy. The consistency and quality of data across multiple years is critical, yet challenging to verify exhaustively.

- **Model Assumptions:** Machine learning models used in this study are based on certain assumptions about data patterns and distributions. Unexpected changes in market behavior, such as regulatory shifts or new market participants, may limit the applicability of these models.

- **External Dependencies:** Since FCR-N prices are influenced by cross-border electricity flows and other external factors beyond Fingrid's control, the models may not fully account for external economic, political, or climatic conditions that influence price trends.

- **Computational Resources:** Some machine learning models, especially those with high complexity, require substantial computational resources. This study balances model accuracy with resource efficiency, but computational constraints may still limit the ability to fully explore all potential models and configurations.

## 1.4 Research Plan and Structure

The following points explain the steps and timeline for completing this thesis:

- **Data Collection and Preprocessing:** Access and preprocess data from Fingrid's open platform, ensuring data consistency and handling any missing values or anomalies.

- **Feature Engineering and Correlation Analysis:** Identify relevant features based on domain knowledge and correlation analysis. Generate lagged features and combinations of variables such as renewable generation and load forecasts.

- **Model Development and Tuning:** Implement multiple machine learning models, such as linear regression, random forest, LSTM etc. Optimize model parameters and select the best-performing models based on accuracy metrics.

- **Model Evaluation and Analysis:** Assess model accuracy using test data, evaluate feature importance, and analyze results to understand FCR-N price influences and model performance.

- **Thesis Writing and Finalization:** Document the research process, analyze findings, and complete each chapter of the thesis. Finalize formatting and revisions based on feedback.

# 2 Literature Review

The accurate forecasting of electricity prices, particularly in reserve markets, has become increasingly important as power systems evolve to accommodate more renewable energy sources. This literature review examines existing research on price forecasting in electricity markets, with a focus on reserve markets and the application of machine learning techniques.

## 2.1 Reserve Market Price Forecasting

Reserve markets play a crucial role in maintaining grid stability, yet they have received less attention in forecasting literature compared to day-ahead and intra-day markets. Olsson and Soder (2008) conducted one of the early studies on forecasting balancing market prices in Northern Europe, highlighting the unique challenges posed by these markets. Their work emphasized the importance of considering the interdependencies between different reserve products and the impact of renewable energy integration on price volatility.

Building on this foundation, Kiesel and Paraschiv (2017) investigated short-term forecasting of balancing power prices in the German electricity market. Their study revealed the significance of incorporating both fundamental market data and historical price patterns to improve forecasting accuracy. However, their model did not specifically address the Frequency Containment Reserve for Normal operation (FCR-N) market, which is the focus of the present study.

Considering the challenges in forecasting reserve market prices, Kraft, Keles, and Fichtner (2020) addresses the challenge of forecasting prices for electricity balancing reserve power, focusing on frequency containment reserve (FCR) prices, which are crucial for optimizing trading positions in competitive auctions. Motivated by a gap in literature on balancing reserve price forecasting, the authors developed a framework combining econometric (SARIMAX) and artificial intelligence (neural networks) approaches, utilizing autoregressive and exogenous factors. Rolling one-step forecasts with model reestimation were implemented. Key findings revealed that while naive forecasts were adequate, neural networks provided superior forecast accuracy over econometric models, although advanced neural configurations offered limited improvement. Econometric models, however, had strengths in explaining price drivers. The study contributes valuable insights for market participants aiming to refine bidding strategies through reliable FCR price forecasting methods.

Furthermore, the paper by Tschora, Pierre, Plantevit, and Robardet (2022) was motivated by the high volatility of electricity prices in the European market and the importance of accurate price prediction for intelligent electricity use. The authors investigate various machine learning techniques for electricity price forecasting, extending current approaches by incorporating previously unused predictive features such as price histories of neighboring countries. Their key findings demonstrate that these additional features significantly improve forecast quality, even during periods of sudden changes. The authors apply a rigorous, transparent, and reproducible methodology, evaluating their models across three European areas over two separate test periods. They show that machine learning models can accurately forecast recent electricity prices and provide valuable information on market dynamics through explainable AI methods.

Huang, Shen, Chen, and Chen (2021) introduces a hybrid model that combines Convolutional Neural Networks (CNN) for feature extraction and Long Short-Term Memory (LSTM) networks for forecasting. This approach outperforms traditional models by effectively capturing both spatial and temporal patterns in volatile electricity price data, showing resilience across various time scales and market conditions. The model handles complex, non-linear price trends better than conventional methods, leveraging CNN for feature detection and LSTM for time-dependent forecasting. This work enhances short-term price forecasting accuracy, offering a flexible and robust tool for dynamic electricity markets.

## 2.2 Machine Learning Approaches in Electricity Price Forecasting

The application of machine learning techniques to electricity price forecasting has gained significant traction in recent years. Machine learning algorithms, such as support vector machines (SVM), decision trees, and ensemble models like random forests, have demonstrated success in energy price forecasting due to their ability to handle complex feature interactions. Weron (2014) provided a comprehensive review of various machine learning techniques used in power market price forecasting, demonstrating their potential to capture complex market dynamics. This work laid the groundwork for more advanced applications of artificial intelligence in energy market forecasting.

Recent research has increasingly focused on the use of deep learning models, particularly recurrent neural networks (RNNs) and their variations, such as long short-term memory (LSTM) networks. These models are designed for sequential data and excel in capturing dependencies in time series data, making them particularly useful in forecasting applications where historical data strongly influence future prices. Lago, De Ridder, and De Schutter (2018) further expanded on this by applying deep learning techniques to forecast electricity prices. Their study showed that deep learning models could outperform traditional statistical methods, particularly in markets with high volatility and non-linear price behaviors. However, the specific application of these techniques to FCR-N price forecasting remains an unexplored area.

On the other hand, Yang, Sun, Hao, and Wang (2022) presents a novel machine learning approach that improves electricity price forecasting accuracy in volatile markets. By combining multiple algorithms with an adaptive model selection strategy, the proposed method outperforms traditional models and hybrid approaches, demonstrating resilience across various markets and volatility levels. The model also highlights influential features for market analysis, offering a flexible and precise tool for modern electricity forecasting needs.

Also, J. Zhang, Tan, and Wei (2020) presents a new hybrid approach designed to improve forecasting accuracy in complex electricity markets. The model combines variational mode decomposition (VMD), self-adaptive particle swarm optimization (SAPSO), SARIMA, and a deep belief network (DBN) to capture both linear and non-linear patterns in price data. This adaptive model outperforms traditional and hybrid methods, showing resilience across time scales and volatile market conditions. It also identifies key factors impacting prices, supporting market analysis and decision-making. This work advances electricity price forecasting by offering a more flexible, robust solution for modern energy markets.

Furthermore, the paper by Meng et al. (2022) presents an advanced approach for electricity price forecasting in markets heavily influenced by renewable energy. By integrating an attention-based LSTM network with crisscross optimization, the model achieves superior accuracy compared to traditional and other deep learning models. Key strengths include the attention mechanism's ability to identify complex relationships affecting price fluctuations and the crisscross optimization's contribution to training efficiency. This model is particularly robust under conditions of high volatility, providing valuable insights into the factors driving prices in renewable-rich markets.

## 2.3 Data-Driven Approaches and Open Data Utilization

The increasing availability of open data in the energy sector has created new opportunities for improving forecasting models. Billé, Gianfreda, Del Grosso, and Ravazzolo (2023) explored the use of open data for electricity price forecasting, demonstrating how publicly available information could enhance prediction accuracy. Their work underscores the potential value of leveraging open data platforms, such as the one provided by Fingrid, in developing more robust forecasting models.

Gabrielli, Wüthrich, Blume, and Sansavini (2022) introduces a data-driven model using Fourier analysis to improve long-term electricity price forecasting with high time resolution. Their approach decomposes price dynamics into base components and volatility elements, achieving accurate long-term predictions by focusing on dominant and residual frequencies. The model effectively generalizes to previously unseen markets and handles data uncertainty via Monte Carlo simulations. This study advances electricity price forecasting by offering a high-resolution, data-driven solution for capturing complex market dynamics over extended periods.

Furthermore, Demir, Mincev, Kok, and Paterakis (2021) explores methods to enhance electricity price forecasting accuracy in situations with limited historical data. The authors investigate data augmentation techniques—such as transformations, autoencoders, and adversarial networks—to generate additional data that supports model training. Key findings indicate that adversarial networks are particularly effective, yielding synthetic data that enhances forecasting accuracy and model robustness. This study highlights the potential of data augmentation to improve model performance and generalization across different market scenarios, providing insights applicable to time series regression beyond electricity pricing.

The hybrid model proposed by R. Zhang, Li, and Ma (2020) improves forecasting accuracy in volatile electricity markets by combining Convolutional Neural Networks (CNN) for feature extraction with Long Short-Term Memory (LSTM) networks for modeling temporal dependencies, effectively handling non-linear and complex patterns in electricity price data. Key findings in their work indicate that this CNN-LSTM framework outperforms both traditional and hybrid models in accuracy and robustness across various time scales and market conditions. The study highlights the potential of integrating deep learning architectures to tackle the unique challenges of day-ahead price forecasting.

## 2.4 Nordic Market Specific Studies

Research specific to the Nordic power market provides valuable insights for FCR-N price forecasting. Kristiansen (2014) examined price forecasting and optimization in the Nordic power market, highlighting the unique characteristics of this market, including the impact of hydropower dominance and cross-border trading. While this study focused on the general Nordic market, it provides a crucial context for understanding the dynamics that may influence FCR-N prices in Finland.

Schütz Roungkvist, Enevoldsen, and Xydis (2020) addresses the challenges of forecasting electricity prices in markets characterized by high penetration of renewable energy, specifically in the Danish electricity market (DK1), where nearly 49% of power production relies on wind energy. The authors develop a forecasting model that incorporates key factors such as electricity consumption, thermal power production, wind generation, and historical electricity prices. The results demonstrate that this linear model is effective in estimating long-term electricity prices in environments with significant wind energy contributions. The findings provide valuable insights for utilities and asset owners in developing risk management strategies and valuing their assets. By emphasizing the importance of accounting for renewable energy production in price forecasting models, the research contributes to the field by presenting a straightforward yet effective approach to addressing the complexities posed by the stochastic nature of renewable energy in electricity markets. Overall, the study highlights the necessity for accurate forecasting methods as markets increasingly integrate renewable energy sources.

The paper by Mehrdoust, Noorani, and Belhaouari (2023) focuses on developing an optimized artificial neural network (ANN) model for forecasting monthly electricity spot prices in the Nordic region. Utilizing a genetic algorithm (GA), the authors optimize the weights and biases of the ANN, comparing its performance against other prediction models such as the Schwartz-Smith stochastic model, Levenberg–Marquardt trained ANN, and particle swarm optimization algorithms. The results indicate that the GA-optimized ANN model significantly outperforms the other methods in terms of prediction accuracy. This innovative model offers a valuable technique for electricity spot price forecasting, assisting energy companies in making informed decisions and effectively managing risks. The authors based their analysis on historical data of monthly electricity spot prices from the Nord Pool market, which is the primary electricity market in the Nordic region, with prices measured in Euros per megawatt-hour (EUR/MWh).

## 2.5 Gap in Current Research

Despite the growing body of literature on electricity price forecasting, there remains a significant gap in research specifically addressing FCR-N price forecasting in the Finnish market using open data. The unique characteristics of the FCR-N market, combined with the availability of Fingrid's open data platform, present an opportunity to develop more accurate and tailored forecasting models for Balancing Service Providers (BSPs).

This thesis aims to address this gap by developing an enhanced price forecasting model for the FCR-N market, leveraging the latest advancements in machine learning techniques and the rich dataset provided by Fingrid's open data platform. By doing so, it seeks to contribute to the broader field of energy market forecasting and provide practical tools for BSPs operating in the Finnish electricity market.

# 3 Dataset

## 3.1 Data Collection

The data used in this study was obtained from Fingrid's open data platform, a comprehensive source for Finnish electricity market data, including forecasts and historical records relevant to energy production and reserve markets. The datasets were selected and collected based on their relevance to Frequency Containment Reserve for Normal operation (FCR-N) price forecasting. The datasets include both generation forecasts for renewable energy sources and consumption forecasts, as well as actual hourly FCR-N prices. These datasets were chosen due to their impact on FCR-N pricing and availability in time intervals conducive to machine learning models that depend on granular data.

The datasets used are as follows:

- **Solar Power Generation Forecast (15-min interval):** This dataset provides a forecast of solar power generation in Finland, updated daily. With data in 15-minute intervals, it allows for detailed insights into solar power's contribution to electricity generation and its influence on reserve prices.
- **Wind Power Generation Forecast (15-min interval):** Similar to solar power forecasts, this dataset includes wind power generation predictions in Finland, also updated daily. Wind generation data, particularly at a 15-minute interval, provides critical information for understanding variability in renewable generation.
- **Electricity Consumption Forecast (15-min interval):** The electricity consumption forecast dataset provides predictive data on Finland's power demand. Accurate demand forecasting is essential for reserve markets, as imbalances between demand and generation impact the reserve requirements and thus influence reserve prices.

- **Frequency Containment Reserve for Normal Operation, Hourly Market Prices (1-hour interval):** This dataset includes hourly market prices for FCR-N, the main target variable of this study. It records historical prices at an hourly interval and serves as a reference for model development and forecasting accuracy.

Data were collected for the period January 1, 2023, 00:00 to August 31, 2024, 23:55, providing a robust sample across different seasons and market conditions. To ensure uniformity in data frequency, the 15-minute interval data (solar and wind power forecasts) were resampled to hourly intervals using mean values, aligning them with the hourly frequency of the FCR-N prices. For better computability and usage, the 15 minutes interval data were resampled to 1 hour interval using the mean values of the original data.

## 3.2 Use of Forecasted Values for Consumption, Solar, and Wind

In this study, forecasted values for electricity consumption, solar power generation, and wind power generation have been utilized instead of actual values. This choice aligns with the practical conditions under which Balance Service Providers (BSPs) operate in the reserve power markets, particularly in forecasting Frequency Containment Reserve for Normal operation (FCR-N) prices. BSPs are required to submit their bids and forecast FCR-N prices at least one day in advance. Since BSPs do not have access to the actual real-time values for these factors during bidding, they rely solely on forecasted data to make predictions.

Using forecasted values ensures that the model realistically mirrors the information BSPs would have at their disposal, allowing for more accurate simulations of real-world forecasting scenarios. By incorporating forecasted consumption, solar, and wind data into the model, this study provides a practical approach to handling the uncertainties that BSPs face in the day-ahead market and accounts for the limitations inherent to the forecasting process. This approach highlights the importance of day-ahead forecasts in price prediction accuracy and in strategic decision-making for BSPs operating in volatile energy markets.

## 3.3 Data Visualization

To gain an initial understanding of the collected data and assess patterns, trends, and seasonal variability, visualizations were created for each dataset. These visualizations provide insight into the temporal behaviors of electricity consumption, renewable power generation, and FCR-N price fluctuations over the study period. Key data visualizations include:

**Time Series Plot of FCR-N Prices**

A time series plot of FCR-N prices across the collected period reveals trends and volatility in reserve prices, highlighting periods of higher reserve costs and potential impacts of seasonal demand shifts.

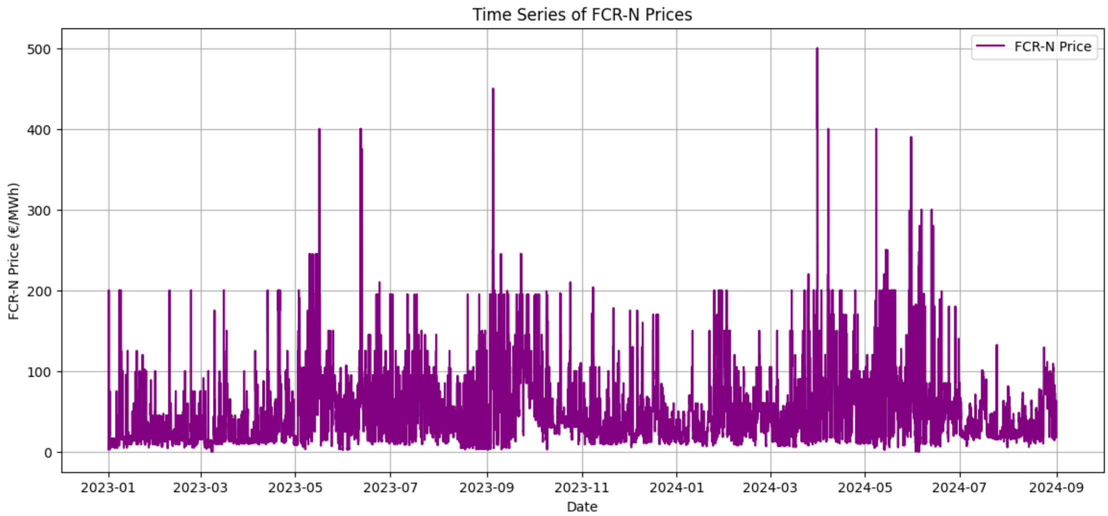


**Figure 1.** Time series plot for FCR-N prices.

Analysis of the FCR-N price data shows an average of 49.5 €/MWh over the specified period. However, the data also reveals a peak price of 500 €/MWh, suggesting that extreme market conditions can lead to substantial price surges. This variability underscores the challenges in forecasting FCR-N prices accurately and the importance of accounting for such fluctuations in reserve markets.

**Electricity Consumption Forecast**

The consumption forecast was visualized to show daily and weekly demand cycles, offering insight into recurring patterns in electricity demand that influence FCR-N pricing.

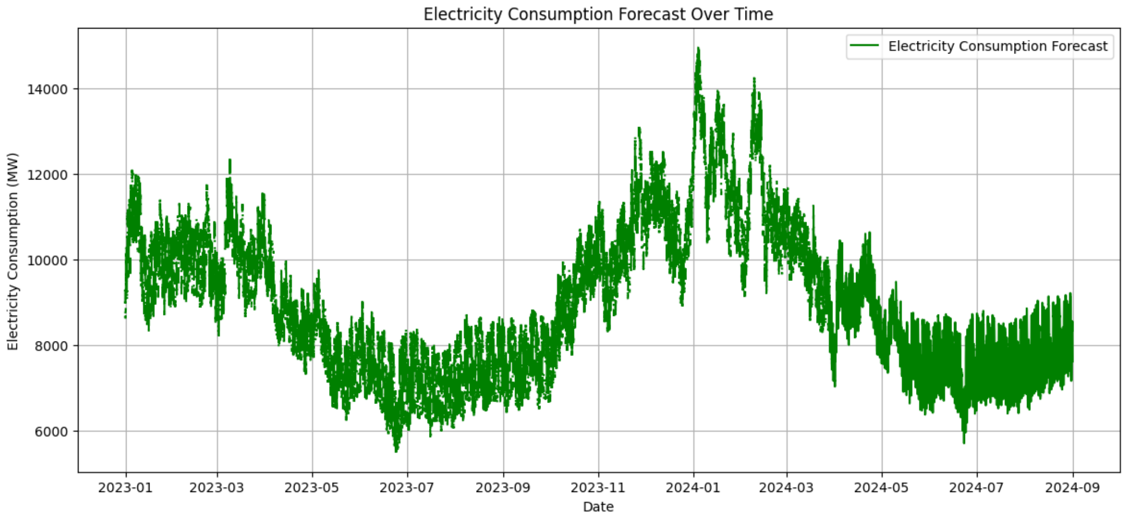


**Figure 2.** Time series plot for Electricity consumption forecast.

The electricity consumption forecast data over the study period reveals an average consumption of 9101.2 MW. The highest consumption level, recorded at 14,957.7 MW, occurred in January 2024, reflecting the typically increased demand for electricity during winter months. In contrast, the lowest forecasted consumption was 5520.6 MW, observed in July 2023. Interestingly, data trends indicate that July tends to be a low-demand month, with both July 2023 and July 2024 showing the lowest levels of electricity consumption. This seasonal pattern aligns with lower industrial and heating demand during summer, underscoring the influence of seasonal variability on electricity consumption forecasts.

**Wind and Solar Power Generation Forecasts**

Separate time series plots for solar and wind generation forecasts were created to observe fluctuations in renewable energy supply. Variability in these sources is expected due to factors such as weather and seasonality, which influence grid demand for reserves.

The wind power generation forecast over the analyzed period shows substantial variabil-

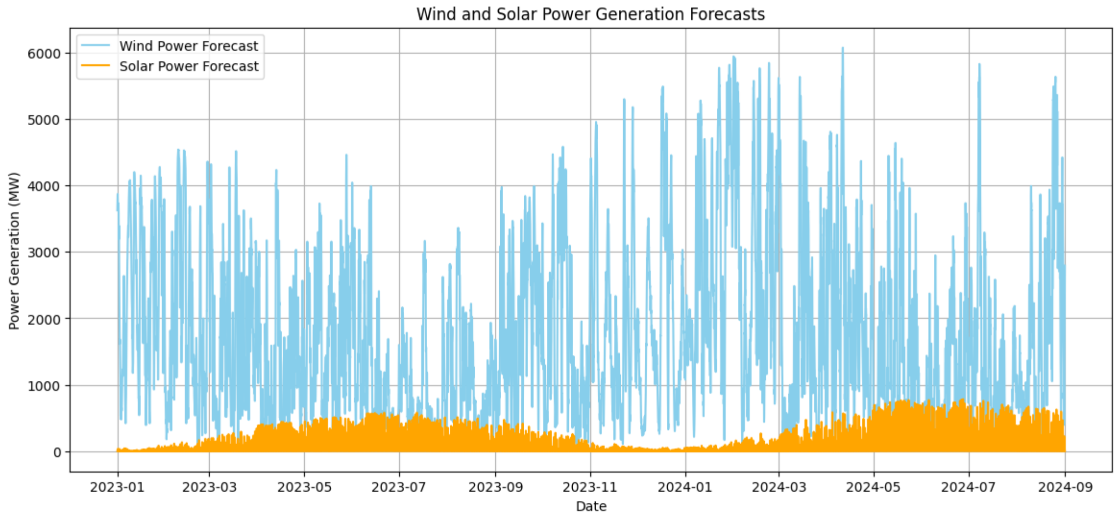


**Figure 3.** Time series plot for Wind and Solar power generation forecast.

ity, with an average generation of 1735.8 MW. The minimum forecasted generation was 26.3 MW, while the maximum reached 6072.6 MW in April 2024, a period marked by higher wind activity. This variability highlights wind power's dependency on seasonal and atmospheric conditions, with periods of heightened generation aligning with specific seasonal patterns. In comparison, the solar power generation forecast has a lower average output of 113.2 MW. Forecasts range from a minimum of 0 MW during winter months, when sunlight is scarce, to a maximum of 782.8 MW, typically observed in the summer. This seasonal trend in solar generation emphasizes the significant influence of daylight hours and seasonal solar intensity on renewable energy contributions.

Each visualization offers a preliminary understanding of how each feature might correlate with or impact FCR-N prices, establishing a foundation for further exploratory data analysis and feature selection in model development.

# 4 Methodology

## 4.1 Data Preprocessing

This sub-section describes the preprocessing steps applied to prepare the dataset for model development. Key preprocessing tasks include resampling, merging, defining calculated features, and performing correlation analysis to select relevant features for our forecasting model. Finally, scaling and splitting the data ensure optimal input for training and evaluation.

### 4.1.1 Resampling

Since some datasets, such as the solar and wind power generation forecasts, were recorded at 15-minute intervals, these were resampled to a one-hour interval using the mean values. This resampling was necessary to align with the hourly interval of the FCR-N price data and ensure consistency across the dataset for effective merging and analysis.

### 4.1.2 Merging the datasets

After resampling, the individual datasets were merged to form a single unified pandas DataFrame (McKinney et al. (2010)). This combined dataset maintains the appropriate values for each feature at the specified hourly intervals. Aligning these time series facilitates a cohesive analysis and ensures all relevant features are available for the corresponding time stamps.

### 4.1.3 Defining non-solar and non-wind generation

This derived feature represents the remaining generation after subtracting the wind and solar contributions and it is crucial for forecasting reserve market prices, as it captures the demand fluctuations not met by renewable energy. This feature allows us to understand the influence of non-intermittent generation dynamics on FCR-N prices more effectively.

#### 4.1.4 Calculating correlations

In this step, we calculated two types of correlations: autocorrelation of the FCR-N price with itself at various time lags, and cross-correlations between FCR-N prices and other variables, including non-solar and non-wind generation.

**Autocorrelation Analysis:** Autocorrelation measures the similarity between observations of a variable at different time points. According to Marco Taboga (2020), it is very helpful in identifying patterns or trends in data and is often calculated for different time lags to examine the relationship between observations at various intervals. In this study, the autocorrelation measures the relationship of FCR-N price with its previous values (lags). Specifically, we analyzed lags from 24 to 48 hours, as we do not have the actual values of FCR-N prices for the current day (lags 0–23). Analyzing these later lags allows us to capture the influence of past FCR-N prices from the prior day and beyond, which can be predictive of future values.

Figure 4 shows that, only specific hours have autocorrelations greater or equal to the threshold (0.3), such that: 24, 25, 26,26, 28, 47, 48.

The correlation plots 5 and 6 illustrate the relationship between non-solar and non-wind generation and the FCR-N (Frequency Containment Reserve - Normal) price. The Pearson correlation plot 5, indicates a negative linear relationship of -0.37 between non-solar and non-wind generation and FCR-N price, suggesting a moderate inverse relationship; as non-solar and non-wind generation increases, FCR-N prices tend to decrease, though the correlation is not very strong. The Spearman correlation plot 6, shows a slightly stronger negative association with a correlation coefficient of -0.44, indicating that there may be a more pronounced monotonic relationship than a purely linear one. This stronger

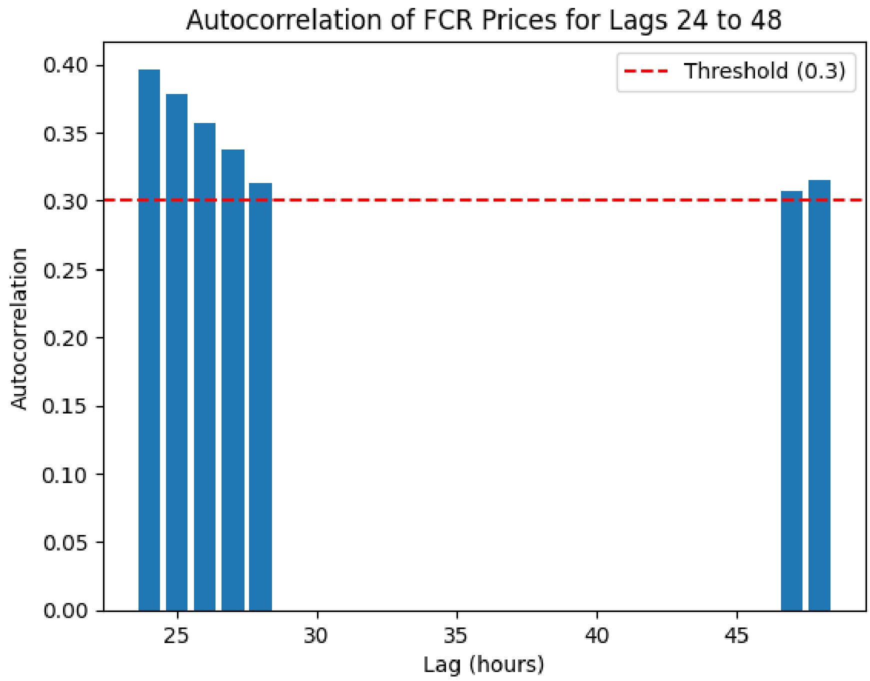


**Figure 4.** Autocorrelation for FCR-N prices for lags 24-48.

Spearman correlation suggests that the rank order of non-solar and non-wind generation and FCR-N price are inversely related, potentially reflecting underlying non-linear trends. Overall, these findings suggest that fluctuations in non-renewable generation capacity moderately affect FCR-N prices, though other factors likely play a role in the price variations as well.

**Cross-Correlations:** We calculated Pearson and Spearman correlations between FCR-N prices and other features (Electricity Consumption Forecast, Wind Power Forecast, Solar Power Forecast, and non-solar and non-wind generation) to assess their relationships. While both correlations provide insight into linear and monotonic relationships respectively, we focused on those with a correlation threshold above 0.3 to identify significant features.

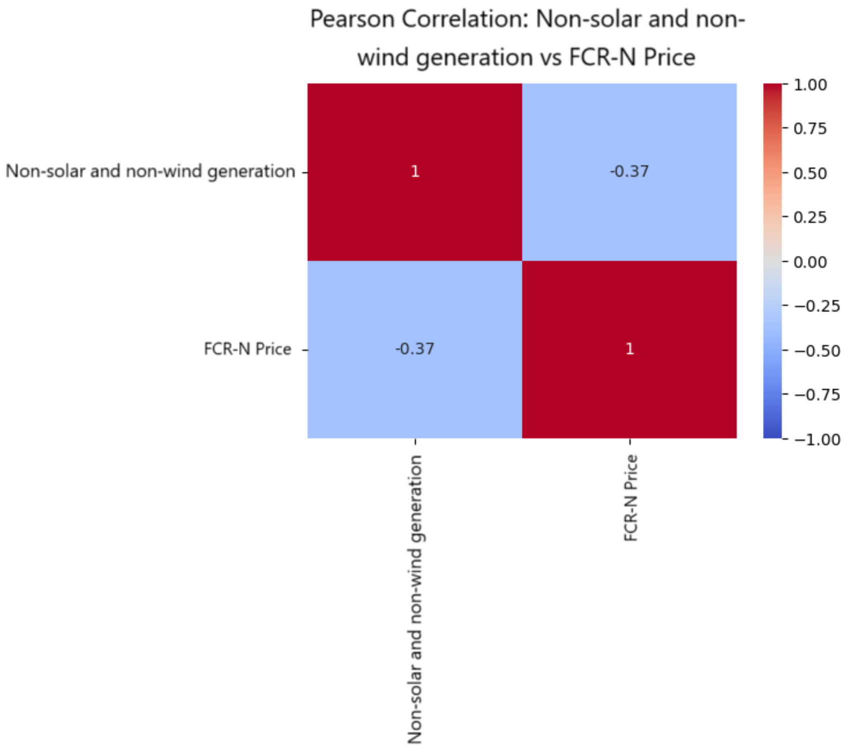


**Figure 5.** Pearson correlation for non-solar and non-wind generation and FCR-N prices.

#### 4.1.5 Getting the final data

Based on the correlation analysis, we observed that the correlations between wind and solar power forecasts with FCR-N prices were below the 0.3 threshold, suggesting a weak relationship. Therefore, these features were excluded from the final dataset to streamline the model and focus on more relevant predictors like non-solar and non-wind generation and the identified significant lags of FCR-N prices.

#### 4.1.6 Scaling the data

Scaling is an essential preprocessing step that standardizes the feature values, improving model performance by ensuring that all features contribute equally. It transforms data by subtracting the feature's minimum value and dividing by the feature's range (maximum - minimum), ensuring the minimum value becomes 0 and the maximum becomes 1. This scaler brings features to a common scale without distorting their ranges, making

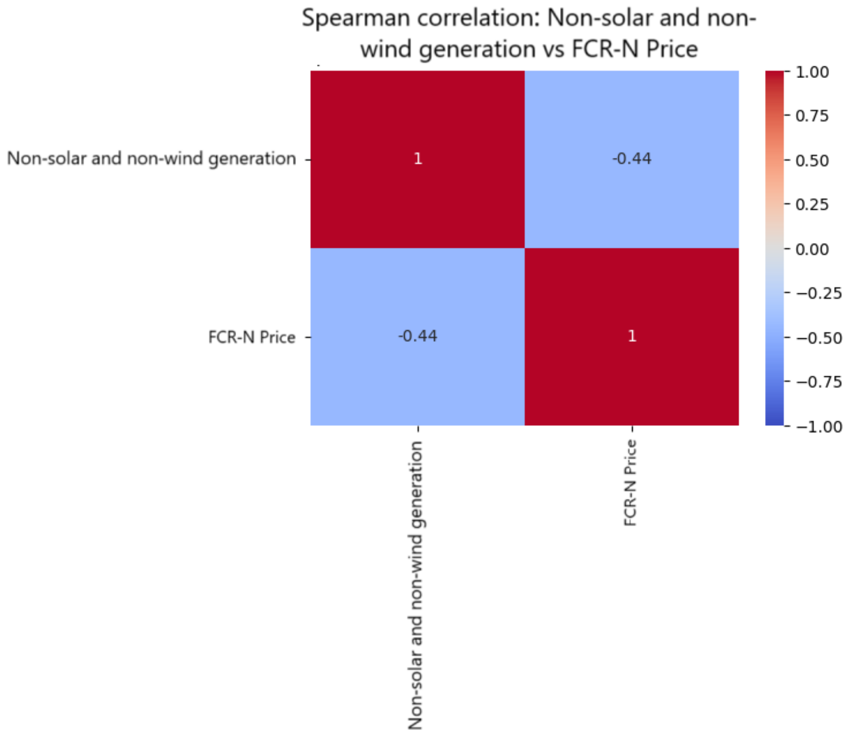


**Figure 6.** Spearman correlation for non-solar and non-wind generation and FCR-N prices.

it suitable for algorithms sensitive to input scale, such as neural networks and those using distance measures. It also preserves zero values, which is beneficial for sparse data, and can handle outliers more effectively than some standardization techniques. MinMax scaling is particularly useful when data distributions are not Gaussian or their distributions are unknown according to Jason Chong (2020). We applied Min-Max scaling in the project, which transforms the values to a range between 0 and 1.

#### 4.1.7 Splitting the data for training

To evaluate model performance effectively, we split the dataset into training and testing sets. This splitting allows the model to be trained on a portion of the data (80%) and then tested on unseen data (20%) to validate its forecasting accuracy. The separation helps prevent overfitting, ensuring the model can generalize well to new data.

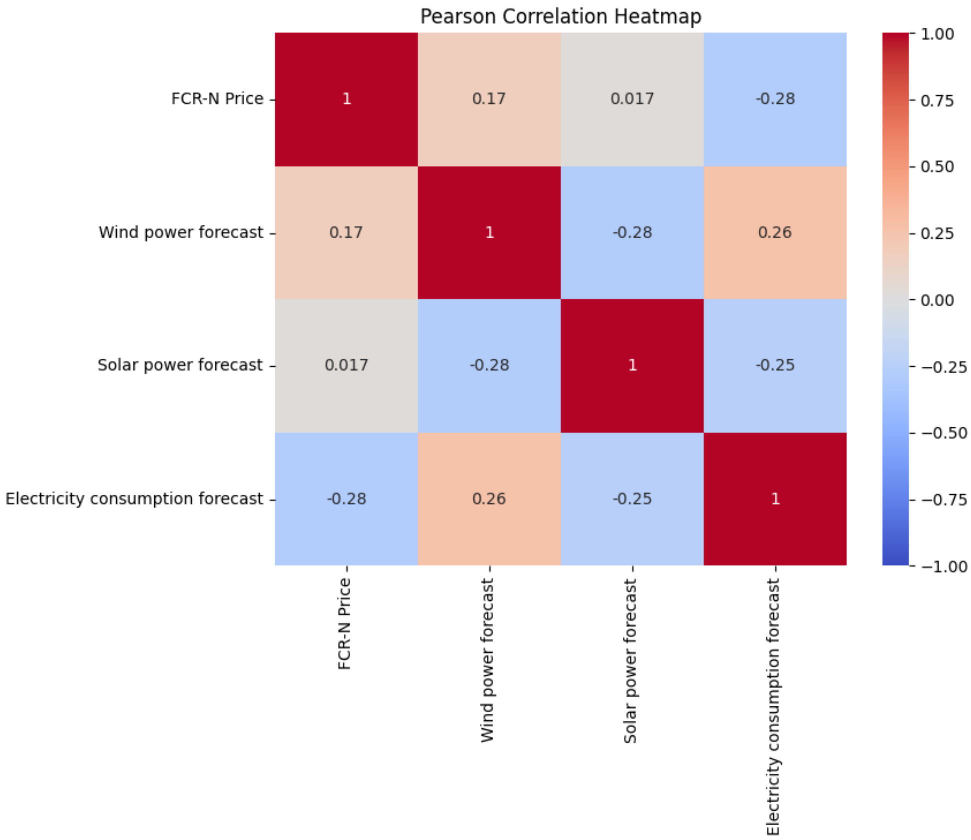


**Figure 7.** Cross-Correlations (Pearson).

## 4.2 Machine Learning Techniques

In this study, we explored a range of machine learning models to forecast FCR-N prices, divided into three main categories: gradient boosting models, regression-based models, and neural network-based models. Each category offers unique strengths suited to different aspects of forecasting.

### 4.2.1 Regression based Models

Regression-based models are traditional machine learning techniques used to model relationships between features and a target variable. They work best for linear relationships but can also handle moderate nonlinear relationships with appropriate adjustments. Here are the regression models applied in this study:

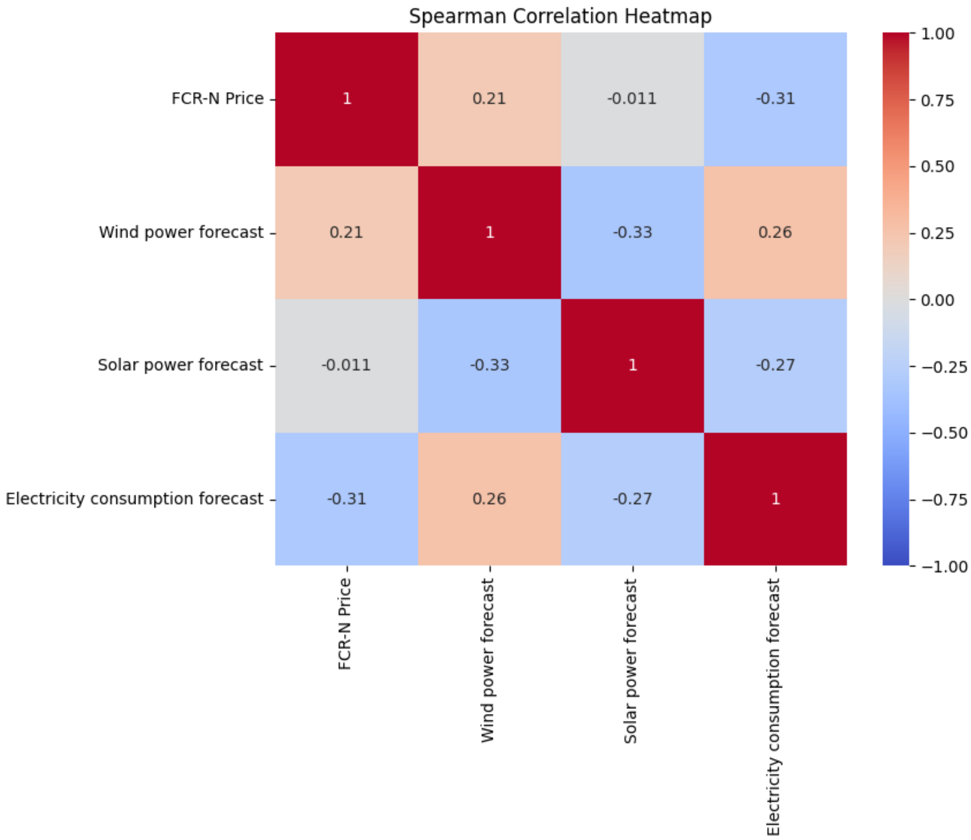


**Figure 8.** Cross-Correlations (Spearman).

- **LinearRegression:** The foundational model in regression analysis, it assumes a linear relationship between the input features and the target variable (Montgomery, Peck, and Vining (2021)). While simple, it can be a strong baseline model.

- **Lasso (Least Absolute Shrinkage and Selection Operator):** This regularized regression technique adds an L1 penalty, shrinking some coefficients to zero, effectively performing feature selection. It is useful when there are many features, as it helps eliminate irrelevant variables (Muthukrishnan and Rohini (2016)).

- **Ridge Regression:** A regression method that applies an L2 penalty to control for large coefficients, making the model more robust to multicollinearity and overfitting (Marquardt and Snee (1975)). Unlike Lasso, Ridge regression reduces the impact of irrelevant features rather than eliminating them.

- **ElasticNet:** Combining L1 and L2 penalties, ElasticNet balances feature selection with regularization. It is beneficial when dealing with highly correlated features or datasets with both sparse and dense features (De Mol, De Vito, and Rosasco (2009)).

- **Polynomial Regression:** Polynomial regression transforms input features into polynomial terms to capture non-linear relationships between features and the target variable (Heiberger, Neuwirth, Heiberger, and Neuwirth (2009)). This model can fit a wider range of data patterns than simple linear regression by adjusting the polynomial degree, though it can be prone to overfitting with high degrees.

- **KNeighborsRegressor:** This non-parametric regression model uses the average of the nearest neighbors' target values for prediction. It works well for small datasets and cases where the target variable depends on similar neighboring values (Imam Muhajir (2019)).

### 4.2.2 Gradient Boosting based Models

Gradient boosting is an ensemble learning technique that builds a strong predictive model by combining multiple weak learners, typically decision trees. These models incrementally learn from previous errors, optimizing performance iteratively. Gradient boosting is particularly effective for complex, nonlinear relationships in data. In this study, we utilized the following models:

- **RandomForestRegressor:** According to El Mrabet, Sugunaraj, Ranganathan, and Abhyankar (2022), it is a bagging technique that builds an ensemble of decision trees trained on different samples of data and averages their predictions. While not technically a boosting algorithm, it complements gradient boosting by enhancing the diversity of predictions.

- **GradientBoostingRegressor:** This model builds sequential trees, each correcting errors from previous trees, resulting in improved accuracy. Gradient boosting is well-suited for continuous data and large datasets (Tomonori Masui (2022)).

- **XGBRegressor (XGBoost):** An optimized gradient boosting model using regularization and parallel processing (Jason Brownlee (2021)). XGBoost's regularization feature makes it robust against overfitting, and it is particularly effective with large, sparse datasets.

- **LGBMRegressor (LightGBM):** According to shilpamahq2o (2024), it is a gradient boosting framework that grows trees leaf-wise rather than level-wise. LightGBM is optimized for speed and memory usage, making it highly efficient for large datasets. It performs particularly well with categorical and high-dimensional data.

### 4.2.3 Neural Network based Models

Neural network models are powerful for capturing complex, nonlinear relationships and sequential dependencies in time series data. They are particularly useful for forecasting, as they can capture temporal patterns and trends. The following neural network models were employed:

- **Recurrent Neural Network (RNN):** RNNs are designed for sequential data, with connections that allow information to persist over sequences. Each output is influenced by previous states, making it effective for time-dependent data. However, standard RNNs are limited by issues such as vanishing gradients in longer sequences (Sherstinsky (2020)).

- **Long Short-Term Memory (LSTM):** LSTM is a variant of RNN designed to capture long-term dependencies by introducing memory cells and gating mechanisms, which allow it to remember information over longer time steps. LSTMs are highly effective for time series forecasting, especially when sequences exhibit long-term dependencies (Yu, Si, Hu, and Zhang (2019)).

- **Gated Recurrent Unit (GRU):** GRU is a simpler alternative to LSTM, with fewer parameters. It has a gating mechanism similar to LSTM but combines the forget and input gates into a single update gate, making it more efficient for smaller datasets and less complex sequences. GRUs are particularly useful for real-time forecasting, as they are computationally efficient (Dey and Salem (2017)).

# 5 Result Analysis

In this section, we evaluate the performance of the different machine learning models applied to forecast the FCR-N prices, analyzing their effectiveness in capturing trends and making accurate predictions. By comparing results from gradient boosting-based models, regression-based models, and neural network-based models, we can identify the optimal approach for reliable forecasting. We also explore error metrics, correlation analysis, and the significance of the explanatory features.

## 5.1 Model Performance Evaluation:

To assess each model's performance, we utilized metrics such as Root Mean Squared Error (RMSE), providing insight into prediction accuracy and error distribution. The square root of mean squared error provides an efficient error metric in the same unit as the data. RMSE is effective for evaluating models when large errors are particularly undesirable.

**Gradient Boosting Models:** Gradient boosting models like XGBoost and LightGBM showed robust performance with lower error metrics compared to other approaches. LightGBM achieved the lowest RMSE and MAE scores, indicating its ability to efficiently handle the complex patterns in FCR-N price data.

**Regression Models:** Among the regression models, Linear Regression and Ridge Regression performed well due to their ability to capture linear and mildly nonlinear relationships, but their predictive power was limited compared to ensemble models. Lasso Regression provides a bit higher error than the previously mentioned models as they struggled with the nonlinear dependencies in the data. Also, Polynomial Regression model struggles the most with the most error of all models.

**Neural Networks:** RNN-based models, particularly LSTM, achieved competitive results, capturing time dependencies effectively. GRU, with its simpler architecture, offered sim-

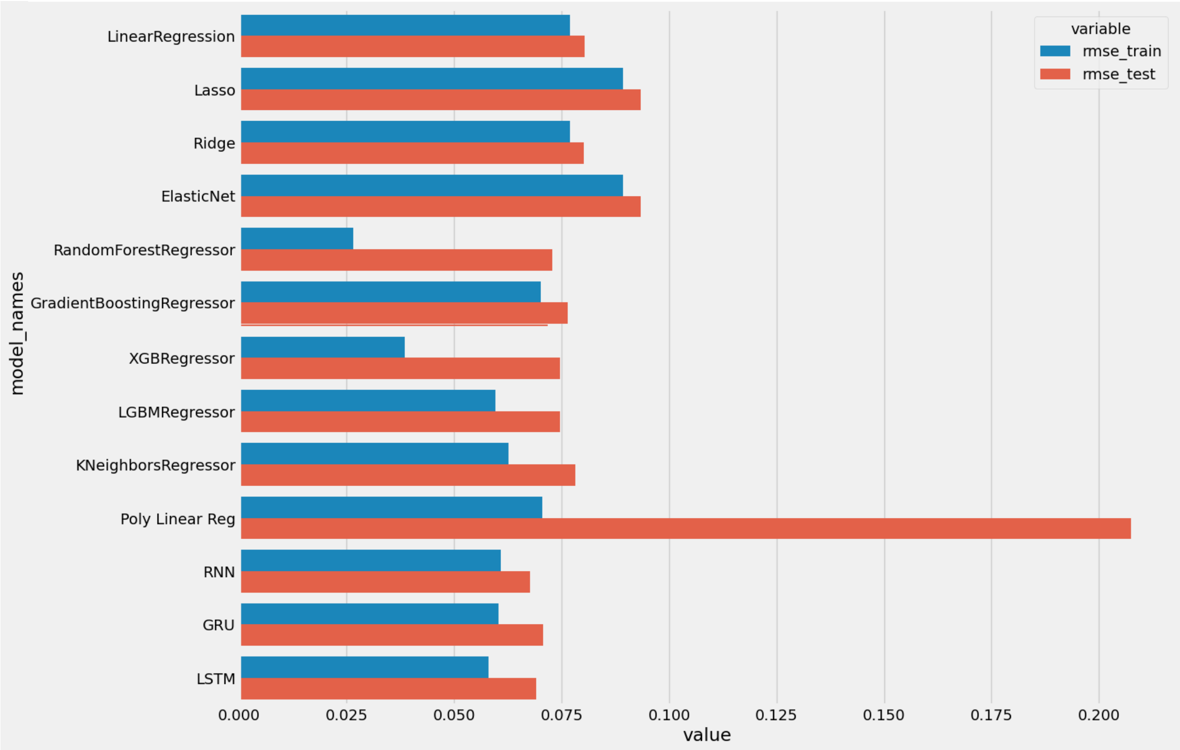


**Figure 9.** Performance comparison of all models.

ilar results with lower computational costs and slightly surpassed LSTM in accuracy.

To summarize, XGBoost, LightGBM, and Random Forest Regressor show the lowest rmse test values, indicating strong generalization ability. These would likely be the best models to choose from for this dataset. Polynomial Linear Regression shows significant overfitting, making it a poor choice. The simple linear models (Linear, Lasso, Ridge, ElasticNet) seem to underfit, with similar and relatively high RMSE for both training and testing sets.

Furthermore, detailed charts were generated that showed comparisons of the predicted data with the real data.

From these charts, we can see that, LSTM has the lease error but both RNN and GRU based models perform similar to the GRU model.

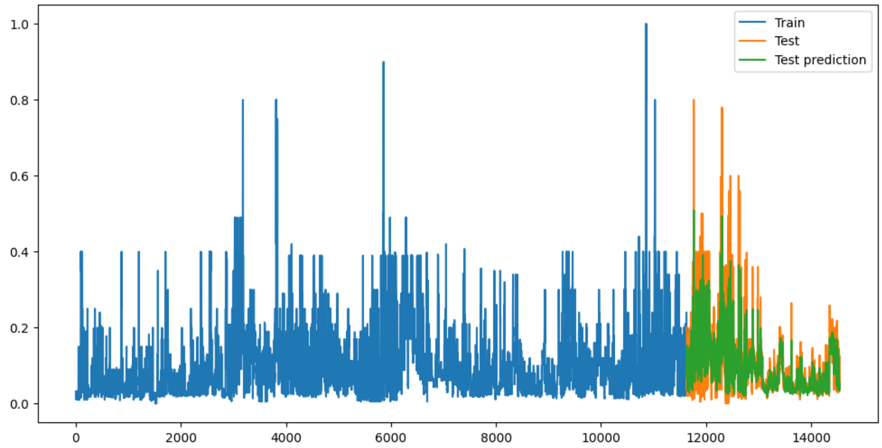


**Figure 10.** The training values of FCR-N prices and the RNN predictions of FCR-N prices on the test set results..

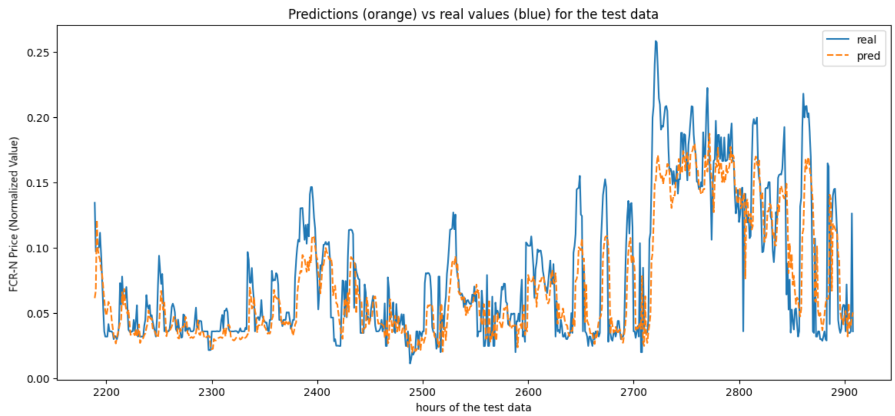


**Figure 11.** Comparison of RNN model predictions (orange) with actual FCR-N prices (blue).

## 5.2 Comparative Analysis of Forecast Accuracy

A comparison of the models revealed the following insights:

**Error Trends:** Gradient boosting models consistently had the lowest error values, suggesting that ensemble methods are well-suited for complex, high-dimensional datasets like FCR-N price data.

**Correlation Analysis:** We observed that the models performed better when lags 24-48 of the FCR-N price were included as features, supporting our hypothesis that the 24-hour interval contributes to capturing daily cyclical patterns in prices. This analysis validated

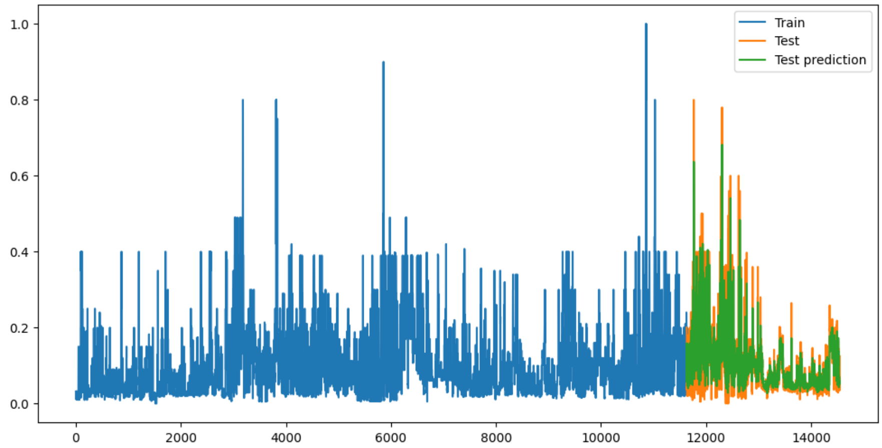


**Figure 12.** The training values of FCR-N prices and the LSTM predictions of FCR-N prices on the test set results..

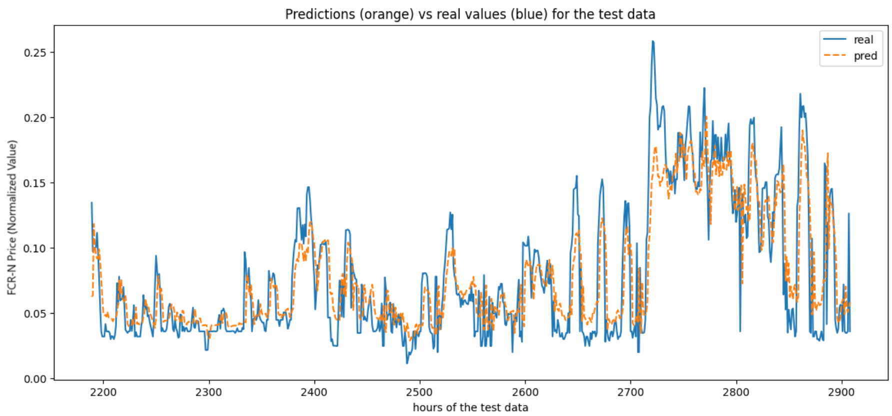


**Figure 13.** Comparison of LSTM model predictions (orange) with actual FCR-N prices (blue).

the exclusion of wind and solar generation forecasts due to low correlation, as they did not significantly improve model accuracy.

**Feature Importance:** The inclusion of non-solar and wind generation and the relevant lags of the FCR-N price (lags 24-48) played a significant role in model accuracy. The correlation analysis showed that non-solar and non-wind generation exhibited a moderate correlation with FCR-N prices, highlighting the dependency of reserve prices on electricity consumption dynamics.

While wind and solar forecasts were initially considered, their correlation with the FCR-N price was low (<0.3), leading to their exclusion in the final dataset. This exclusion

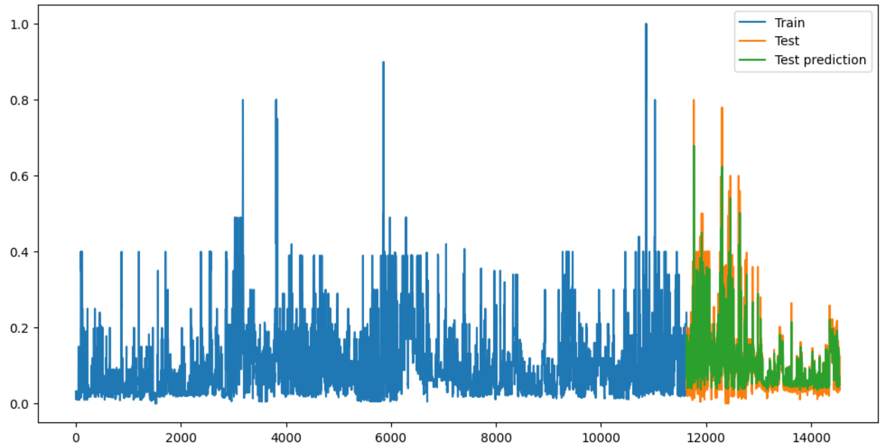


**Figure 14.** The training values of FCR-N prices and the GRU predictions of FCR-N prices on the test set results..

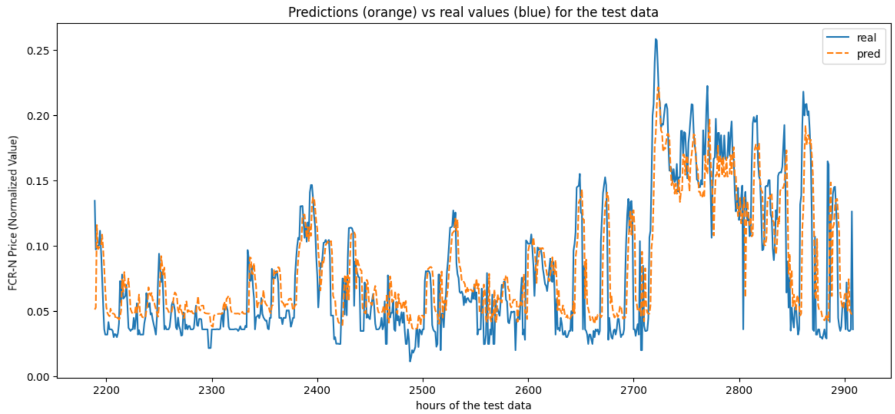


**Figure 15.** Comparison of GRU model predictions (orange) with actual FCR-N prices (blue).

demonstrated a slight improvement in model accuracy, indicating that wind and solar forecasts had minimal predictive power for reserve market prices.

Gradient boosting models allowed us to analyze feature importance. Features such as non-solar and non-wind generation and selected FCR-N price lags (24-48) showed the highest influence on predictions, reaffirming the importance of these factors in price forecasting.

Lastly, the analysis confirms that non-linear models like LightGBM and sequence-based models like LSTM offer significant improvements in forecasting accuracy for FCR-N prices. The sensitivity of FCR-N prices to changes in non-solar and non-wind generation and previous prices underlines the importance of including these variables in the modeling process.

The high performance of gradient boosting models and neural networks also highlights the potential of machine learning techniques for enhancing price forecasting accuracy in reserve markets, aiding BSPs in better aligning their strategies with market dynamics.

# 6 Discussion

## 6.1 Theoretical Implications

This study contributes to the field of energy economics and machine learning-based forecasting by focusing on FCR-N (Frequency Containment Reserve for Normal operation) price predictions within reserve markets. By implementing advanced machine learning models, this research provides insights into the performance and suitability of different models—such as gradient boosting algorithms and neural networks—in handling price fluctuations. The findings support the idea that non-linear and sequence-based methods (like LightGBM and LSTM) are highly effective in predicting reserve market prices, which are subject to complex, dynamic, and seasonally fluctuating patterns.

Additionally, the analysis suggests that reserve market prices are moderately influenced by demand variables like non-solar and non-wind generation but minimally impacted by renewable generation forecasts, particularly wind and solar. This has theoretical significance, highlighting how traditional supply-demand models might need adjustments in contexts where renewable energy integration is high but its forecasting is less relevant for reserve pricing. These insights contribute to refining forecasting methods for energy markets and potentially improving reserve market policies.

## 6.2 Limitations

While this study offers valuable insights, it also has several limitations that could affect the interpretation of results:

**Data Granularity and Availability:** The dataset's reliance on hourly and 15-minute intervals, though relevant, may limit model performance. Forecasting accuracy could benefit from data with higher temporal granularity. Additionally, data availability for specific factors affecting FCR-N prices, such as cross-border market dynamics or extreme weather, was limited and not fully incorporated.

**Model Selection and Complexity:** While various models were employed, each model has inherent limitations. For instance, linear models are restricted by their inability to capture non-linear patterns, while complex models like neural networks require more data for effective training. The performance of models such as LSTM and LightGBM might vary in real-time applications due to these dependencies on training data.

**Exclusion of External Market Factors:** The study did not fully consider external market influences, such as international power trade and demand shifts in neighboring countries. These factors can significantly impact reserve prices, especially in interconnected markets like that of Finland.

**Interpretability of Complex Models:** Models like LSTM and LightGBM, while accurate, are often considered black-box models, making it challenging to interpret the underlying factors driving price fluctuations. This limitation reduces the transparency of forecasting in operational settings.

## 6.3 Future Research Possibilities

Building upon this study, several avenues for future research could further enhance FCR-N price forecasting accuracy and applicability in reserve markets:

**Incorporating External Market Data:** Future research could integrate additional market data, such as cross-border trading volumes, international reserve prices, or market-wide demand patterns. This would provide a holistic view of factors influencing reserve prices and help improve forecasting models.

**Exploring Advanced Deep Learning Architectures:** While this study included basic neural network models, future research could explore advanced deep learning architectures such as transformers or hybrid models combining LSTM and gradient boosting. These approaches may better capture complex temporal patterns, potentially enhancing forecasting accuracy.

**Model Explainability and Interpretability:** Efforts to improve model interpretability, especially for complex models, would make machine learning applications more useful to operators and policymakers. Techniques such as SHAP (SHapley Additive exPlanations) or LIME (Local Interpretable Model-agnostic Explanations) could be applied to shed light on feature importance and model predictions.

**Real-time Forecasting Models:** Research focused on developing real-time or near-real-time forecasting models could be valuable for reserve markets. By training models to handle streaming data, researchers can design algorithms that adapt to incoming information and provide more accurate short-term price forecasts.

**Impact of Renewable Energy Integration:** Given the limited effect of wind and solar forecasts on FCR-N prices observed in this study, further research could investigate under what conditions renewable forecasts significantly impact reserve pricing. As renewable integration continues to grow, this could provide insights into adapting forecasting strategies over time.

In summary, this research highlights the potential of machine learning models in reserve market forecasting, while also identifying areas where future work can build upon these foundations to address the complexities of evolving energy markets.

# 7 Conclusion

This thesis explored advanced machine learning techniques to enhance the forecasting accuracy of Frequency Containment Reserve for Normal operation (FCR-N) prices in reserve markets, using data from Fingrid's open data platform. The primary goal was to address the challenges associated with forecasting FCR-N prices by examining the predictive performance of various machine learning models, including regression-based models, gradient boosting algorithms, and neural networks. Through this approach, the study aimed to provide insights for Balancing Service Providers (BSPs) in effectively navigating the dynamic pricing of reserve markets.

The study's findings highlight that machine learning models, particularly tree-based gradient boosting methods and recurrent neural networks (RNNs), can effectively capture the complex temporal patterns in FCR-N price data. Among the variables analyzed, the non-solar and non-wind generation and historical lags of FCR-N prices proved to be the most significant predictors of future prices, while renewable generation forecasts (wind and solar) showed minimal direct correlation with FCR-N prices. This aligns with the observation that reserve pricing is more influenced by non-intermittent generation requirements rather than direct fluctuations in renewable energy forecasts, due to factors like forecast uncertainty and the priority of reliability in reserve markets.

Despite the promising results, several limitations were encountered, including the data granularity and interpretability of complex models like RNNs. The findings, therefore, suggest opportunities for further research to incorporate additional features, such as international market factors, and to explore more interpretable deep learning techniques. By refining forecasting models with real-time adaptability and interpretability, future research can advance the applicability of machine learning in reserve markets.

In conclusion, this thesis contributes to the growing body of research on energy price forecasting by demonstrating the efficacy of machine learning models in reserve markets.

These models not only provide operational insights for BSPs but also offer a foundation for policymakers and market operators to improve price stability and resource allocation in balancing energy systems. As reserve markets evolve with increasing renewable energy integration and market interconnection, machine learning will play an essential role in advancing data-driven decision-making in energy management.

# Bibliography


Billé, A. G., Gianfreda, A., Del Grosso, F., & Ravazzolo, F. (2023). Forecasting electricity prices with expert, linear, and nonlinear models. *International Journal of Forecasting*, *39*(2), 570-586. https://doi.org/https://doi.org/10.1016/j.ijforecast.2022.01.003

Demir, S., Mincev, K., Kok, K., & Paterakis, N. G. (2021). Data augmentation for time series regression: Applying transformations, autoencoders and adversarial networks to electricity price forecasting. *Applied Energy*, *304*, 117695.

De Mol, C., De Vito, E., & Rosasco, L. (2009). Elastic-net regularization in learning theory. *Journal of Complexity*, *25*(2), 201–230.

Dey, R., & Salem, F. M. (2017). Gate-variants of gated recurrent unit (gru) neural networks. In *2017 ieee 60th international midwest symposium on circuits and systems (mwscas)* (pp. 1597–1600).

El Mrabet, Z., Sugunaraj, N., Ranganathan, P., & Abhyankar, S. (2022). Random forest regressor-based approach for detecting fault location and duration in power systems. *Sensors*, *22*(2), 458.

Gabrielli, P., Wüthrich, M., Blume, S., & Sansavini, G. (2022). Data-driven modeling for long-term electricity price forecasting. *Energy*, *244*, 123107.

Heiberger, R. M., Neuwirth, E., Heiberger, R. M., & Neuwirth, E. (2009). Polynomial regression. *R Through Excel: A Spreadsheet Interface for Statistics, Data Analysis, and Graphics*, 269–284.

Huang, C.-J., Shen, Y., Chen, Y.-H., & Chen, H.-C. (2021). A novel hybrid deep neural network model for short-term electricity price forecasting. *International Journal of Energy Research*, *45*(2), 2511–2532.

Imam Muhajir. (2019). *K-neighbors regression analysis in python.* Retrieved from https://medium.com/analytics-vidhya/k-neighbors-regression-analysis-in-python-61532d56d8e4

Jason Brownlee. (2021). *Xgboost for regression.* Retrieved from https://machinelearningmastery.com/xgboost-for-regression/

Jason Chong. (2020). *What is feature scaling why is it important in machine learning?* Retrieved from https://towardsdatascience.com/what-is-feature-scaling-why-is-it-important-in-machine-learning-2854ae877048

Kiesel, R., & Paraschiv, F. (2017). Econometric analysis of 15-minute intraday electricity prices. *Energy Economics*, *64*, 77-90. https://doi.org/https://doi.org/10.1016/j.eneco.2017.03.002

Kraft, E., Keles, D., & Fichtner, W. (2020). Modeling of frequency containment reserve prices with econometrics and artificial intelligence. *Journal of Forecasting*, *39*(8), 1179–1197.

Kristiansen, T. (2014). A time series spot price forecast model for the nord pool market. *International Journal of Electrical Power Energy Systems*, *61*, 20-26. https://doi.org/https://doi.org/10.1016/j.ijepes.2014.03.007

Lago, J., De Ridder, F., & De Schutter, B. (2018). Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. *Applied Energy*, *221*, 386-405. https://doi.org/https://doi.org/10.1016/j.apenergy.2018.02.069

Marco Taboga. (2020). *Autocorrelation*. Retrieved from https://www.statlect.com/fundamentals-of-statistics/autocorrelation

Marquardt, D. W., & Snee, R. D. (1975). Ridge regression in practice. *The American Statistician*, *29*(1), 3–20.

McKinney, W., et al. (2010). Data structures for statistical computing in python. In *Scipy* (Vol. 445, pp. 51–56).

Mehrdoust, F., Noorani, I., & Belhaouari, S. B. (2023). Forecasting nordic electricity spot price using deep learning networks. *Neural Computing and Applications*, *35*(26), 19169–19185.

Meng, A., Wang, P., Zhai, G., Zeng, C., Chen, S., Yang, X., & Yin, H. (2022). Electricity price forecasting with high penetration of renewable energy using attention-based lstm network trained by crisscross optimization. *Energy*, *254*, 124212.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis*. John Wiley & Sons.

Muthukrishnan, R., & Rohini, R. (2016). Lasso: A feature selection technique in predictive modeling for machine learning. In *2016 ieee international conference on advances in computer applications (icaca)* (pp. 18–20).

Olsson, M., & Soder, L. (2008). Modeling real-time balancing power market prices using combined sarima and markov processes. *IEEE Transactions on Power Systems*, *23*(2), 443-450.

Schütz Roungkvist, J., Enevoldsen, P., & Xydis, G. (2020). High-resolution electricity spot price forecast for the danish power market. *Sustainability*, *12*(10), 4267.

Sherstinsky, A. (2020). Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network. *Physica D: Nonlinear Phenomena*, *404*, 132306.

shilpamahq2o. (2024). *Regression using lightgbm.* Retrieved from https://www.geeksforgeeks.org/regression-using-lightgbm/?itm_source=auth&itm_medium=contributions&itm_campaign=articles

Tomonori Masui. (2022). *All you need to know about gradient boosting algorithm - part 1. regression.* Retrieved from https://towardsdatascience.com/all-you-need-to-know-about-gradient-boosting-algorithm-part-1-regression-2520a34a502

Tschora, L., Pierre, E., Plantevit, M., & Robardet, C. (2022). Electricity price forecasting on the day-ahead market using machine learning. *Applied Energy*, *313*, 118752.

Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, *30*(4), 1030-1081. https://doi.org/https://doi.org/10.1016/j.ijforecast.2014.08.008

Yang, W., Sun, S., Hao, Y., & Wang, S. (2022). A novel machine learning-based electricity price forecasting model based on optimal model selection strategy. *Energy*, *238*, 121989.

Yu, Y., Si, X., Hu, C., & Zhang, J. (2019). A review of recurrent neural networks: Lstm cells and network architectures. *Neural computation*, *31*(7), 1235–1270.

Zhang, J., Tan, Z., & Wei, Y. (2020). An adaptive hybrid model for short term electricity price forecasting. *Applied Energy*, *258*, 114087.

Zhang, R., Li, G., & Ma, Z. (2020). A deep learning based hybrid framework for day-ahead electricity price forecasting. *IEEE Access*, *8*, 143423–143436.