MSc ESDA Title Page

UCL Candidate Code: Group 1 (LSHM1, PKHJ8, MHCR1, LDYY7, PDXT9)

Module Code: BENV0091

Module Title: Energy Data Analysis

Coursework Title: Forecasting Day-ahead prices in NO2 bidding zone using ENTSO-e data

Module Leader: Dr. Amir Gharavi

Date: 09/01/2025

Word Count: 5071

1. Introduction and context

Following the Russian invasion of Ukraine in 2022, European electricity prices have risen significantly and experienced high levels of volatility, on the back of higher natural gas prices (Zakeri et al., 2023).

High and volatile prices can be politically destabilizing, slow economic growth, increase inflation and make it difficult for businesses to budget effectively (Fabra, 2023). It is therefore in the interest of political decision-makers and market participants to understand what factors influence electricity prices in their area and to be able to forecast prices to inform bidding strategies, operational planning and risk management.

This project focuses on the Norwegian NO2 bidding zone, the southern-most bidding zone in Norway.

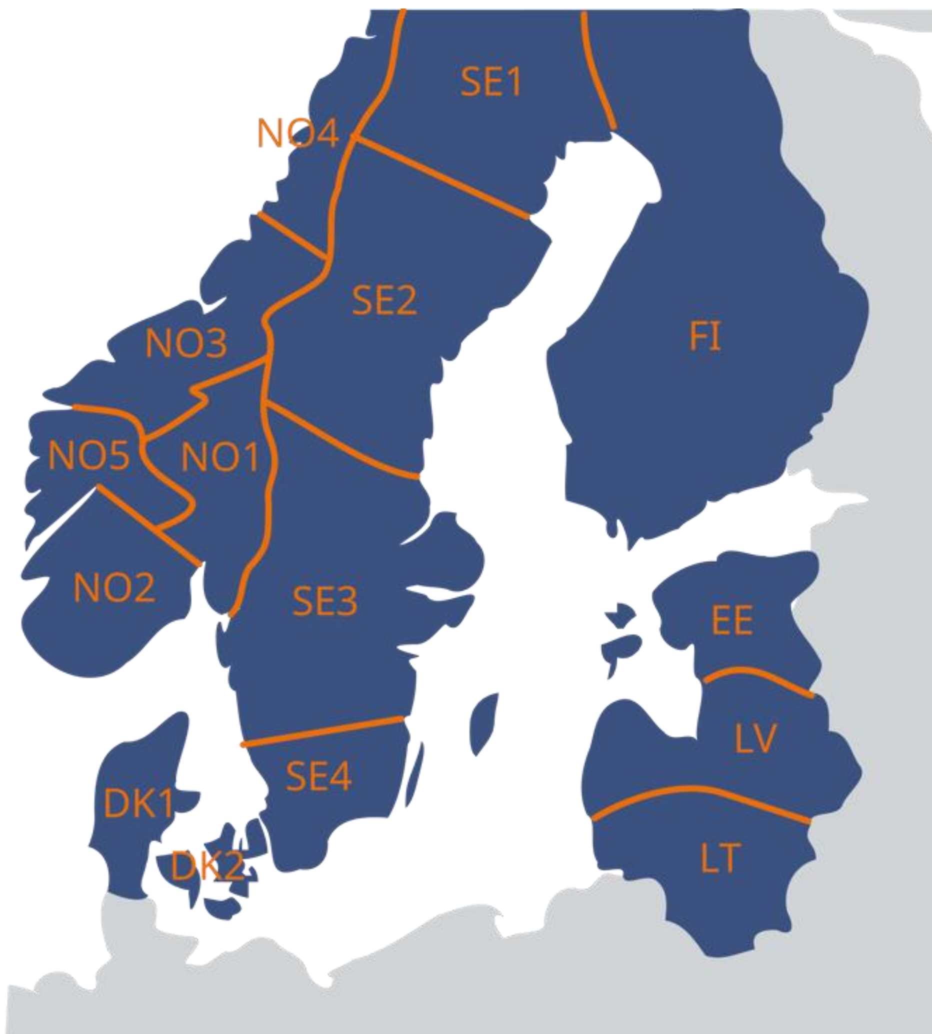

Fig. 1. Nordic bidding zones (Wikipedia Commons, 2022)

The different bidding zones can be visualised in Fig. 1. The NO2 bidding zone is particularly relevant due to ongoing discussions in Norway about whether the zone should cut off its interconnectors to Europe (Richard Milne, 2024). The bidding zone is currently connected to bidding zones in Germany, Denmark, the UK and the Netherlands. Within Norway, NO2 has the highest installed hydropower capacity of any bidding zone, which can be seen in Fig. 2. From Fig 3 it can be seen that nearly all power production in NO2 comes from hydropower.

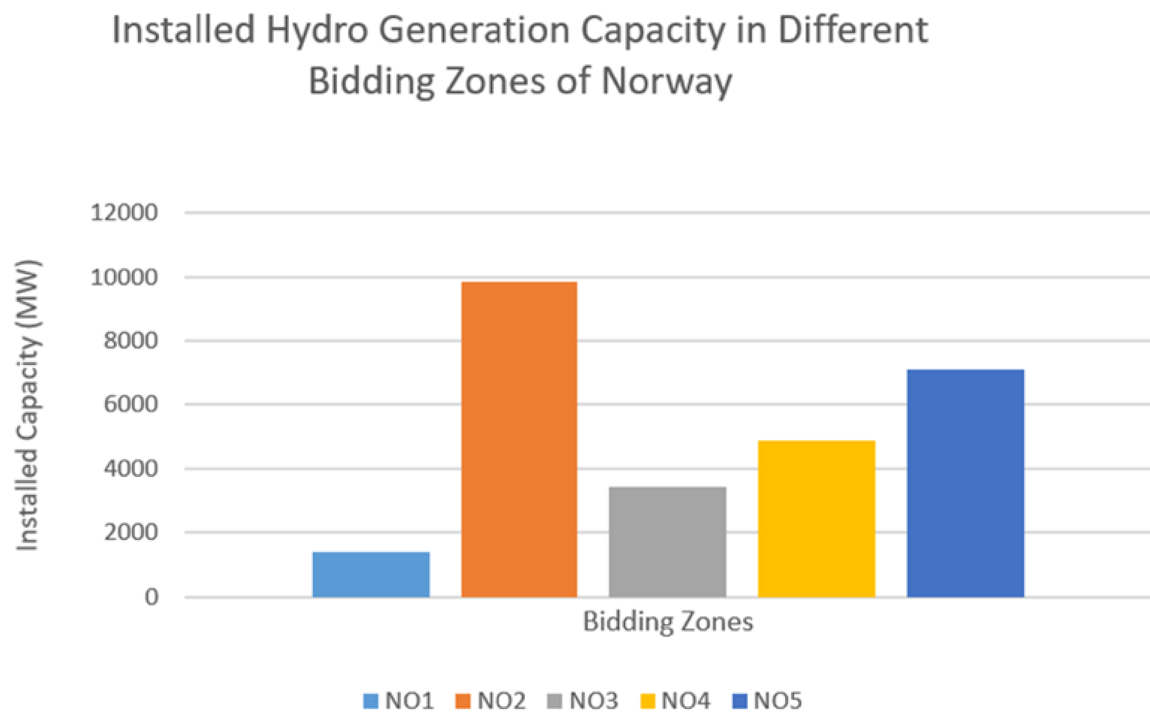

Fig. 2. Installed hydropower capacity across Norwegian bidding zones

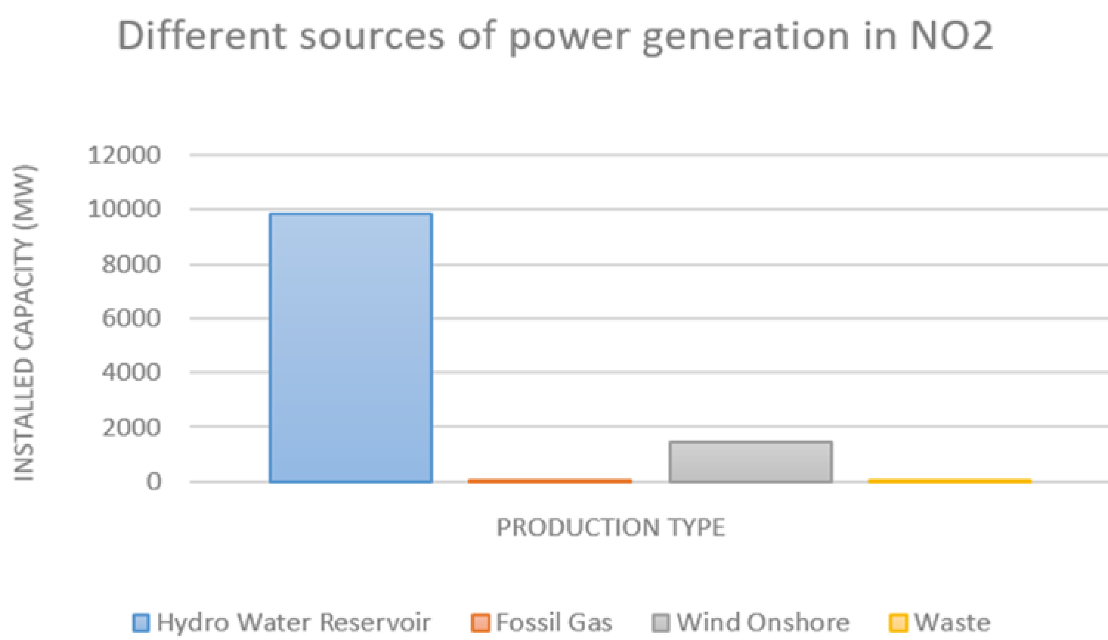

Fig. 3. Power generation sources in NO2 bidding zone (ENTSO-E, 2024b)

Despite the large amount of installed hydropower capacity, Day-ahead power prices in NO2 are also the highest in Norway. As shown in the Fig. 4 displaying aggregate electricity prices, NO2 consistently incurs higher prices, which further emphasises the need for targeted forecasting approaches in this area.

Fig. 4. Aggregate electricity prices across Norwegian bidding zones (NordPool, 2024)

The aim of this project is to predict these NO2 Day-ahead prices using various regression and machine learning models. The performance of each model will then be evaluated against each other.

2. Literature review

Electricity is a unique commodity in that it is difficult to store and supply and demand in the electricity grid must always be exactly matched (Borenstein, 2001). Following the liberalisation of European power markets in the 1990s (Rathke, 2015), these characteristics gave rise to unique price dynamics in the electricity market, which in turn made electricity price forecasting (EFP) an important part of energy companies' decision-making (Bunn, 2004).

Most of the EFP research is dedicated to predicting the Day-ahead price (Jędrzejewski et al, 2022), also known as the spot electricity price. This price is formed on the Day-ahead market, in which market participants submit bids to produce electricity in auctions for each hour of the following day (Weron, 2014). The system operator then determines the market clearing price (MCP) for each hour by accepting bids in order of increasing prices until the total expected load for the following day is met.

Separate auctions take place for each bidding zone in Europe, most of which are defined by national borders (ACER,2024). However, Norway is divided into five different bidding zones to reflect the transmission constraints in the grid between different parts of the country, and the auctions for each of these take place on the NordPool power exchange (Norwegian Ministry of Petroleum and Energy, 2008).

Until the early 2010s, the EFP field was dominated by small linear regression models while recent years have seen a rise in use of machine learning models (Jędrzejewski et al, 2022). Within Europe, most EFP studies have used datasets from large, continental markets such as Germany, France and Belgium (Jędrzejewski et al, 2022).

However, multiple EPF studies have used datasets from Nordic markets and Norway in particular. In many of these studies, the aim has been to find drivers of the Norwegian electricity price, rather than evaluating performance across different models. For example, Bendiksen and Løining (2024) used an ordinary least squares (OLS) regression model to analyse an extensive NO2 dataset, finding a pronounced effect of natural gas prices on electricity prices.

Høyslo and Hjertner (2022) used a multiple linear regression model on an NO2 dataset and found a strong correlation between the Day-ahead price, gas futures and EUA futures.

Viken (2024) utilised neural networks and extreme gradient boosting (XGBoosting) models to illustrate the adverse effect of the energy crisis on EFP model performance.

Common for many EFP studies is the use of extensive datasets with large amounts of features from multiple sources, some of which are always easily accessible (Lago et al, 2021). This leads to issues in evaluating results in the EFP literature and reproducing the research.

To ensure the reproducibility of our findings, this study will only use publicly available data from the European Network of Transmission System Operators for Electricity (ENTSO-e). Whereas much of the literature on NO2 price forecasting has used extensive datasets from multiple sources, this study aims to evaluate which models perform the best on simple supply/demand data from ENTSO-e. These results can then be used to compare with studies using higher-dimensional datasets.

3. Methodology

3.1 Data description and pre-processing

The ENTSO-e Transparency Platform is likely the single most important data source for European electricity markets (Hirth et al, 2018). The platform provides access to power system data on an hourly basis, which can be accessed through a graphical user interface (GUI), file transfer protocol (FTP), or a restful application programming interface (API). We utilized the API service to retrieve datasets directly into our Python environment.

The restful API package comes with two clients, one which returns data in an XML format and one which returns data parsed as Pandas Series or DataFrame. We used the Pandas client for easier processing and converted all Pandas Series to DataFrames.

The response, or dependent variable, in our project is NO2 Day-ahead prices. This is the market clearing price for every hour of the following day in the NO2 bidding zone. The independent variables collected for predicting the price include load forecasts, generation forecasts, wind and solar forecasts, hydro reservoirs, cross-border flows and net transfer capacities. Below is a detailed description of all the features used in the project.

3.2 Feature Description

Day-Ahead Prices

Day-ahead electricity prices for the NO2 zone in Norway represent as the target variable. Additionally, the prices from the previous day ("Prev_Day_DA_prices_NO_2") are included as a predictor to account for price trends.

Load Forecasts

Load forecasts represent the predicted electricity demand in various regions. Forecasted load for NO2 ("Load_forecast_NO_2") is a primary feature, reflecting the anticipated demand in the target region. Forecasts for neighboring areas ("Load_forecast_NO_1", "Load_forecast_NO_5", "Load_forecast_DK", "Load_forecast_NL", and "Load_forecast_DE_LU") are included to account for regional interdependencies.

Renewable Generation Forecasts

Renewable energy sources, particularly wind and solar, theoretically impact electricity prices due to their variability. Forecasted wind ("Wind Onshore_NO_2") and solar generation ("Solar_DK", "Solar_NL") from multiple regions provide insights into the availability of renewable energy in the market.

Generation Forecasts

Generation forecasts indicate the expected electricity production capacity. This includes forecasts for NO2 and other neighboring regions such as "Generation_forecast_NO_2". These features highlight the supply-side dynamics in the electricity market.

Net Transfer Capacities (NTC)

NTC values represent the planned capacity for cross-border electricity transfers. Features such as "NTC_WeekAhead_NO_2_to_NL" capture the transfer capabilities between NO2 and neighboring regions, reflecting market flexibility and trade potential.

Cross-Border Physical Flows

Physical electricity flows between regions are calculated as the net difference between flows in both directions. Features like "Net_Flow_NO_2_to_DE_LU" provide a view of electricity exchanges, which influence supply-demand balances.

Aggregate Water Reservoirs

Norway's electricity market also relies on hydropower. "Aggregate_Water_Reservoirs_NO_2" captures the water reservoir levels, offering insights into hydropower availability, a crucial supply factor in the region.

Feature Engineering

Additional features are derived through lagging, rolling statistics, and interaction terms:

- *Rolling Statistics:* Rolling means and standard deviations over 3, 7, and 30-day windows.
- *Interaction Terms:* Ratios and differences between load and generation forecasts
- *Temporal Features:* Day of the week, month, and weekend indicators to capture seasonality.

3.3 Exploratory data analysis (EDA)

EDA aims to reveal patterns, relationships, and anomalies in the data. This step is crucial for identifying underlying trends, understanding feature behavior, and ensuring that the dataset is ready for modeling. By exploring these aspects, we can better select relevant predictors, design appropriate features, and address potential issues such as multicollinearity or missing data, which might otherwise compromise model performance. The analysis is methodical, beginning with an overview of the dataset and progressing to detailed feature-specific insights.

3.3.1 Data Overview

The dataset starts from October 2023 to September 2024, with hourly data points for all features. The initial inspection revealed missing values in some features, which were addressed using forward-fill and backward-fill methods. Descriptive statistics provided the following insights, which informed preprocessing and modelling decisions. For example, understanding the variability and distribution of features allowed us to identify outliers, address missing values, and determine feature engineering priorities. These insights also guided the selection of features most relevant for capturing the dynamics of day-ahead electricity prices.

3.3.2 Target Variable Analysis

The target variable, "DA_prices_NO_2," exhibited a mean value of approximately 54.36 EUR/MWh and a standard deviation of 29.82 EUR/MWh as shown in Fig 5. The distribution of "DA_prices_NO_2" was analyzed to understand price trends and volatility: A histogram was used to visualize the overall distribution as shown in Fig 6, highlighting price variations and potential outliers.

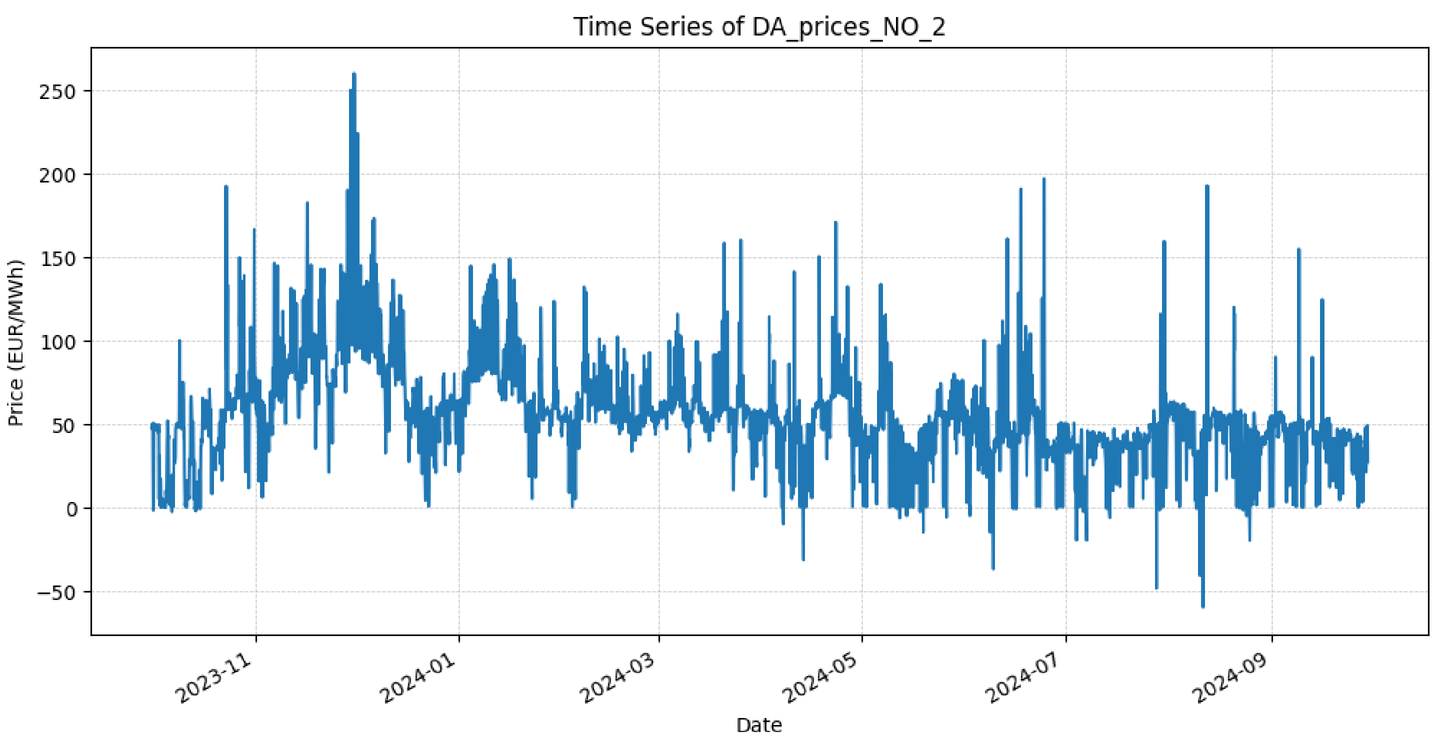

Fig. 5. Historical day-ahead electricity prices of Norway

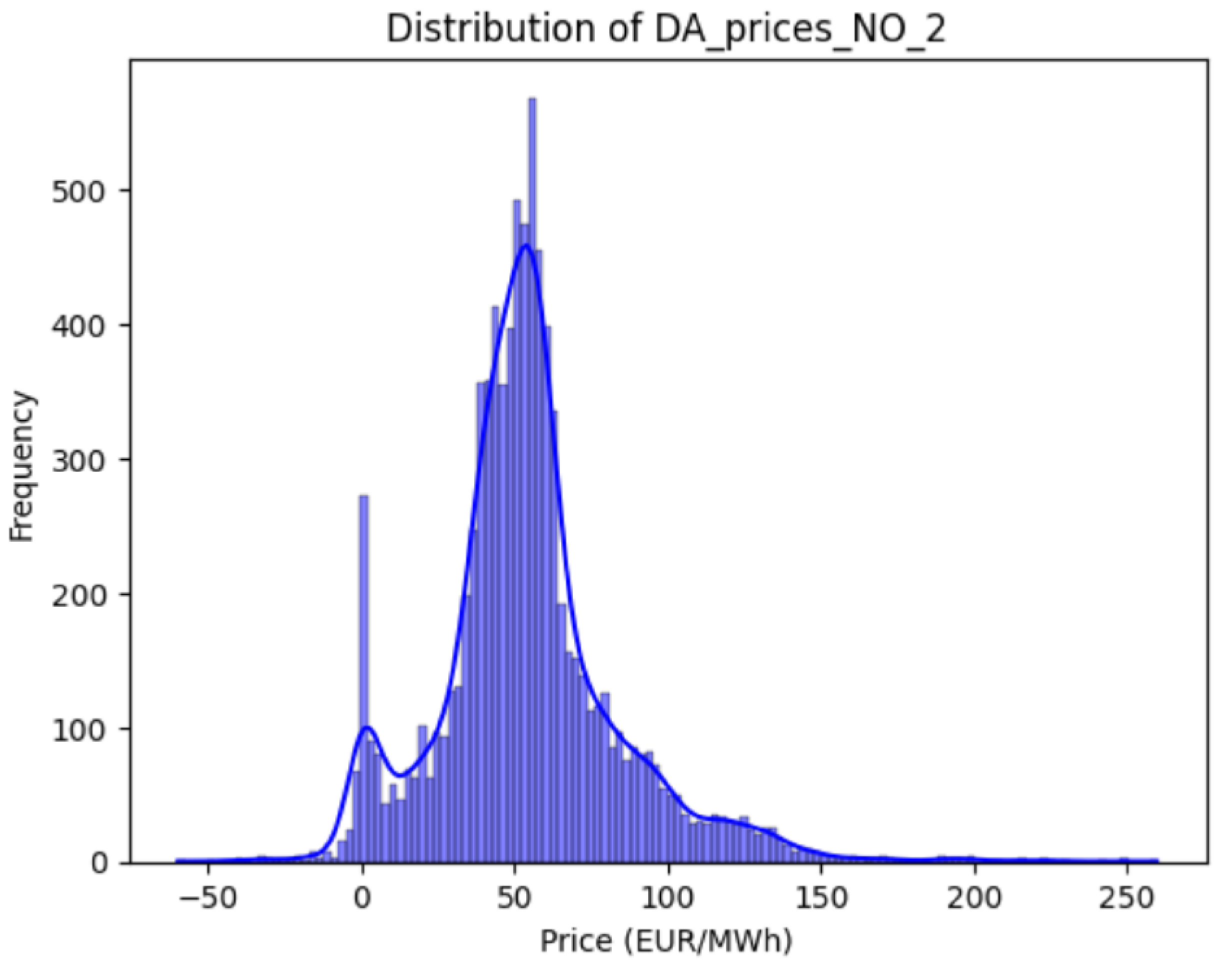

Fig. 6. Histogram of electricity prices of Norway

A histogram revealed a slightly skewed distribution, with occasional price spikes indicative of market stress.

3.3.3 Correlation Analysis

A correlation heatmap highlighted relationships among features:

The feature "Generation_forecast_NO_2" and "Load_forecast_NO_2" showed a strong positive correlation with the target variable, "DA_prices_NO_2" (r = 0.71 and 0.61 respectively), indicating that higher generation and load forecasts are associated with increased electricity prices.

Renewable generation features, such as "Wind Onshore_NO_1" and "Solar_DK," exhibited weak negative correlations with "DA_prices_NO_2," reflecting their tendency to suppress electricity prices due to the increased availability of low-cost renewable energy.

Cross-border flows, represented by "Net_Flow_NO_2_to_NL," showed moderate correlations with "DA_prices_NO_2," highlighting their importance in balancing regional supply and demand dynamics within the electricity market.

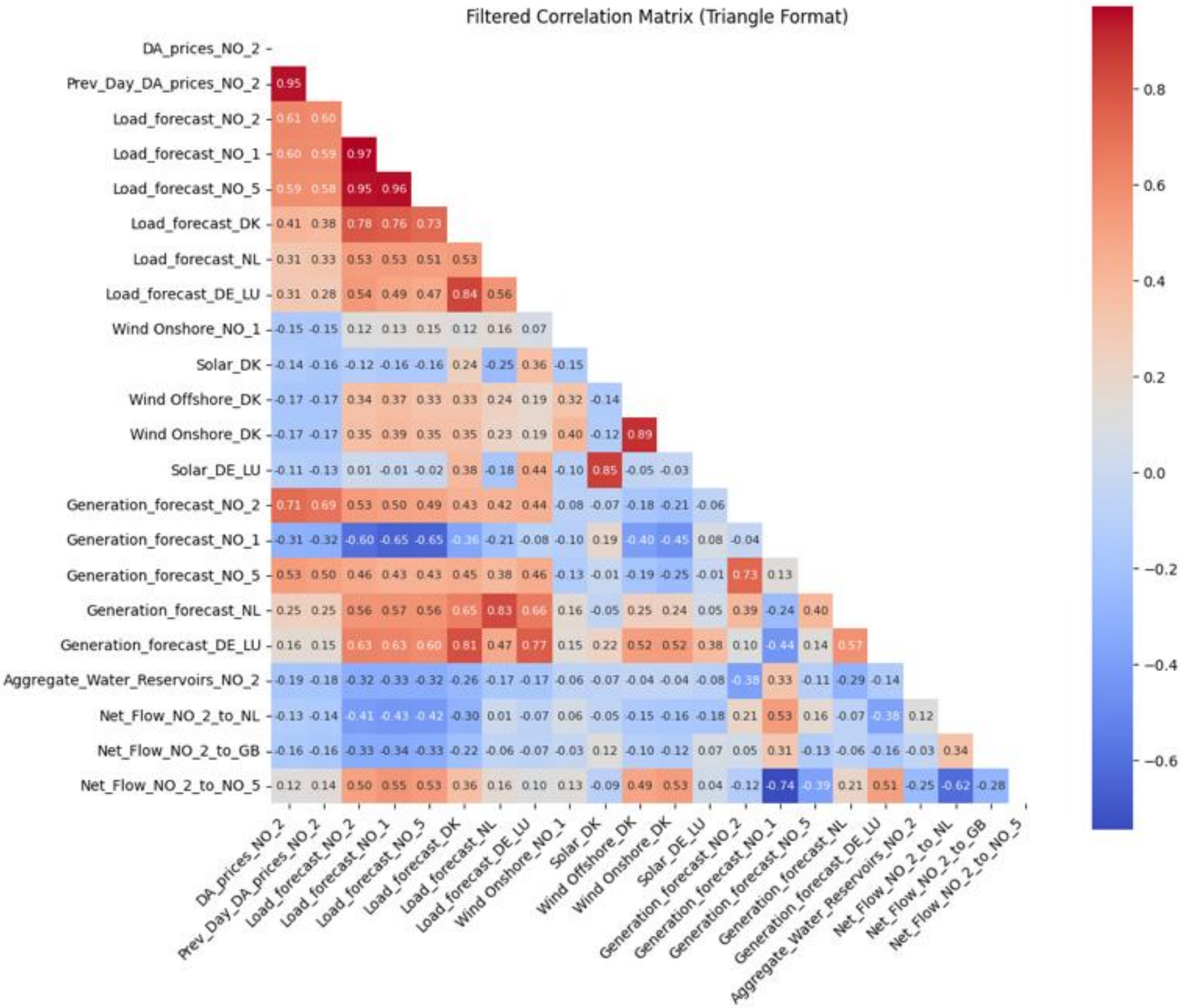

Fig. 7. Feature correlation heatmap

3.3.4. Feature-Specific Analysis

The analysis of load forecasts revealed notable variability, with significantly higher electricity demand observed on weekdays compared to weekends. This could assume that weekdays experience greater energy consumption due to business operations.

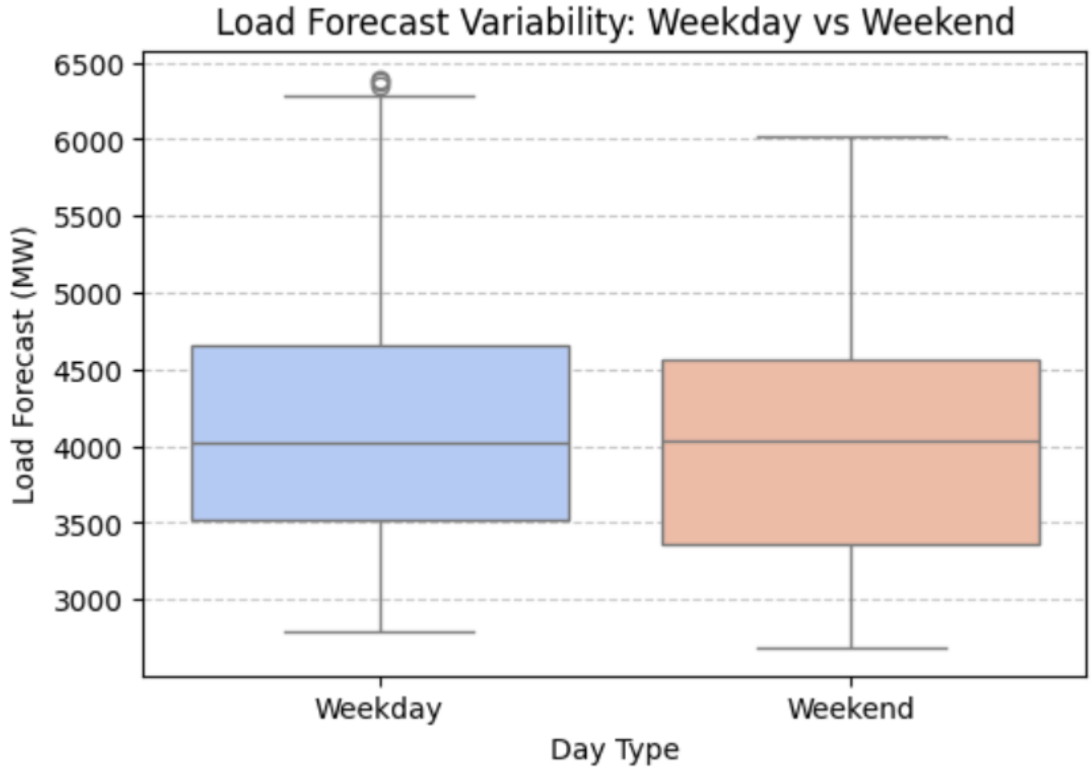

Fig. 8. Load forecast variability

Scatter plots examining the relationship between net transfer capacities (NTC) and day-ahead electricity prices indicated weak linear correlations. This suggests that while NTC values contribute to the overall market dynamics, their direct influence on price formation may be limited.

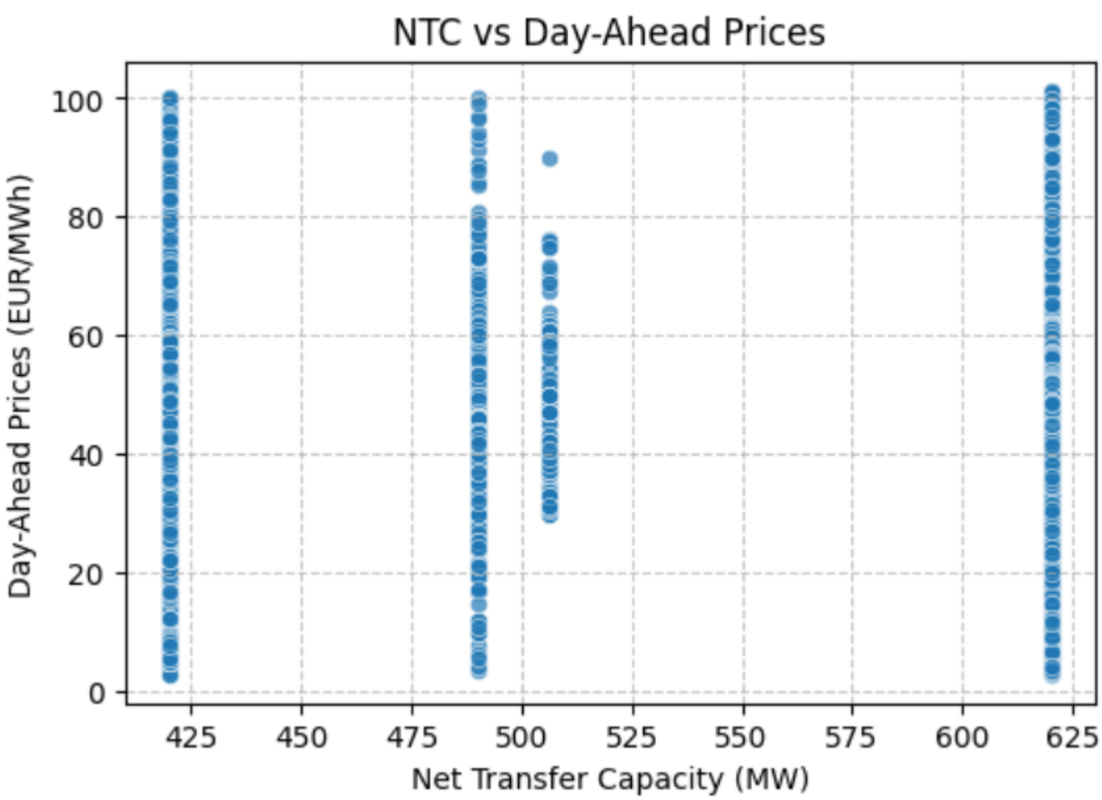

Fig. 9. NTC vs day-ahead price

3.3.5. Outlier Detection

Outliers in the dataset were identified using the Interquartile Range (IQR) method. For the target variable, "DA_prices_NO_2," extreme values were observed, which were likely caused by supply shortages or sudden demand spikes. Similarly, renewable generation features exhibited outliers, particularly during periods of unusually high wind or solar activity. These anomalies were carefully removed from the dataset to ensure the robustness of the subsequent modeling process and to prevent undue influence on the results.

3.3.6. Feature Redundancy

Variance Inflation Factor (VIF) analysis was conducted to assess multicollinearity among the features. High VIF values indicate redundancy, as these features are strongly correlated with others in the dataset. Features with extremely high VIF values (e.g., >20) were flagged.

For example, "Load_forecast_NO_1" (32.5) and "Generation_forecast_DK" (28.0) may be simplified or replaced with aggregated metrics.

3.3.7. Feature Engineering

To improve the model's predictive performance, feature engineering techniques were applied. They focused on capturing temporal patterns, variability, and interactions within the data. Lagged features were created for "Load_forecast_NO_2" and "Generation_forecast_NO_2" at intervals of 1, 7, and 30 days to account for temporal dependencies. These lagged features enable the model to incorporate historical information, which is essential for understanding recurring patterns and delayed effects on electricity prices.

Rolling statistics, including rolling mean and standard deviation, were calculated over 3, 7, and 30-day windows for the same predictors. This step aimed to quantify short and medium-term variability, reflecting periods of stability or volatility in energy demand and supply, which can influence market behavior. Additionally, percentage change features were added to capture relative fluctuations, such as sudden surges in load or generation, which are critical in identifying rapid shifts that may impact day-ahead prices.

Interaction features were used to represent the relationship between load and generation forecasts. The "Load_Generation_ratio" highlights the balance between supply and demand, while the "Load_Generation_diff" quantifies the magnitude of this imbalance. These features are particularly valuable for modeling market dynamics, as they directly relate to price fluctuations during periods of surplus or shortage.

Temporal features were also derived to account for seasonal and weekly trends. The "Day_of_Week" and "Is_Weekend" indicators reflect variations in demand between weekdays and weekends, while the Month feature captures broader seasonal effects, such as higher winter energy consumption.

Finally, the dataset was split into training and testing subsets based on time, and features were standardized using StandardScaler to maintain consistency and numerical stability. These engineered features are expected to enhance the model's ability to identify patterns and predict electricity prices with greater accuracy.

4. Theory & Results

Traditional machine learning methods were evaluated and tested against each other in this study to predict electricity prices.

The visualizations of different models present the actual price predictions of our chosen model for the three-month period from July to October 2024. Our visualization consists of three main parts: actual prices, predicted prices, and confidence intervals.

4.1 Models

4.1.1 Linear regression model

Linear regression is one of the most widely used statistical methods for predictive modelling.

Theory & mathematical foundation

In this study, the dependent variable is the Day-ahead electricity price, while the independent variables could include various factors like load forecasts, renewable generation forecasts, and cross-border flows. The goal of the linear regression model is to find the line of best fit that minimizes the sum of squared differences (errors) between the observed and predicted values. The formula is provided below:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

It is important to apply this model with caution as it can be easy to overfit the data as feature input increases. There are performance metrics which allows for the consideration of this overfitting which will be discussed in more detail in the model accuracy section.

Implementation specifics

In the implementation of predicting NO2 zone prices with linear regression models, we choose to utilize the scikit-learn library as the core modelling framework.

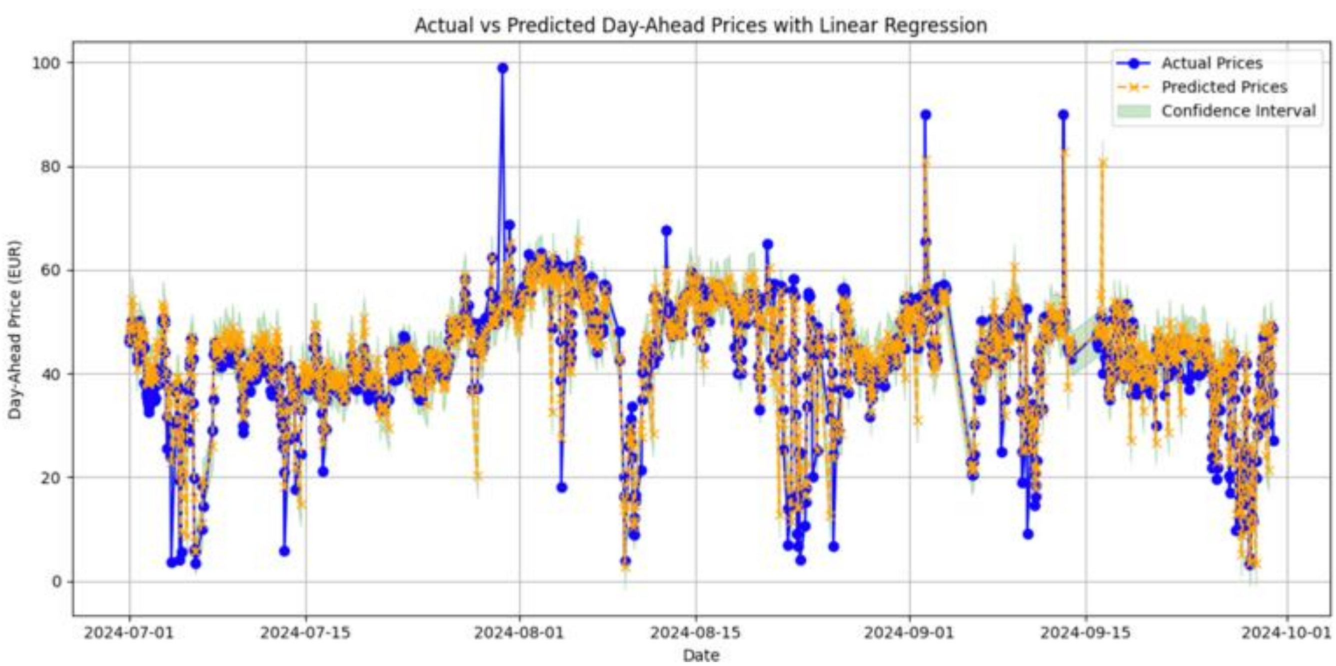

Fig. 10. Actual vs Predicted Day-Ahead Prices with Linear Regression

From Fig. 10 it throughout the observation period, the predicted and actual prices mainly fluctuate between 0 to 60 euros. However, between July and September 2024, there were three distinct price peaks, i.e. real prices exceeding 80 euros. At the end of July 2024, the actual price amounted to 100 euros, but the forecast price did not show such volatility.

```
Top 10 Feature Importances for Linear Regression:
                              Feature  Importance
48        Load_forecast_NO_2_pct_change   18.659620
49  Generation_forecast_NO_2_pct_change    8.252185
0              Prev_Day_DA_prices_NO_2    0.783152
53                                Month    0.383333
30             Load_forecast_NO_2_lag_1    0.007965
37         Load_forecast_NO_2_roll_std_3    0.005101
21               Generation_forecast_DK    0.002596
1                    Load_forecast_NO_2    0.002359
45        Load_forecast_NO_2_roll_std_30    0.002214
18             Generation_forecast_NO_2    0.002113
```

Fig. 11. Feature importance based on Linear Regression

The feature importance scores from Fig. 11 clearly reflect the influence of different factors on electricity prices under the linear regression model, and there is a clear hierarchical structure among the influencing factors. The feature importance score under the linear regression model represents the separate impact of each standardized feature on price prediction.

“Load_forecast_NO_2_pct_change” is the main factor with an importance score of 18.66, followed by “Generation_forecast_NO_2_pct_change” with an importance score of 8.25. Under the linear regression model, "Load_forecast_NO_2_pct_change" dominates the change. The large gap between load forecasts and other characteristics suggests that market participants react more strongly when load forecasts change. This shows that market participants regard demand uncertainty as the main risk factor, and they pay more attention to the demand side.

4.1.2 Decision Tree

Theory & mathematical foundation

The Decision Tree model is a type of supervised machine learning algorithm that is used for both classification and regression tasks. For regression tasks, the Decision Tree tries to minimize the variance of the target variable within each leaf. Decision Trees are often used as they have an advantage of being easy to interpret and understand. However, they can be prone to overfitting, especially when the tree is deep, and the dataset is noisy.

Implementation specifics

We implemented the decision tree model using DecisionTreeRegressor of scikit-learn. This approach allows for a nonlinear relationship between features and "DA_prices_NO_2", making it possible to capture complex market dynamics that linear models may miss.

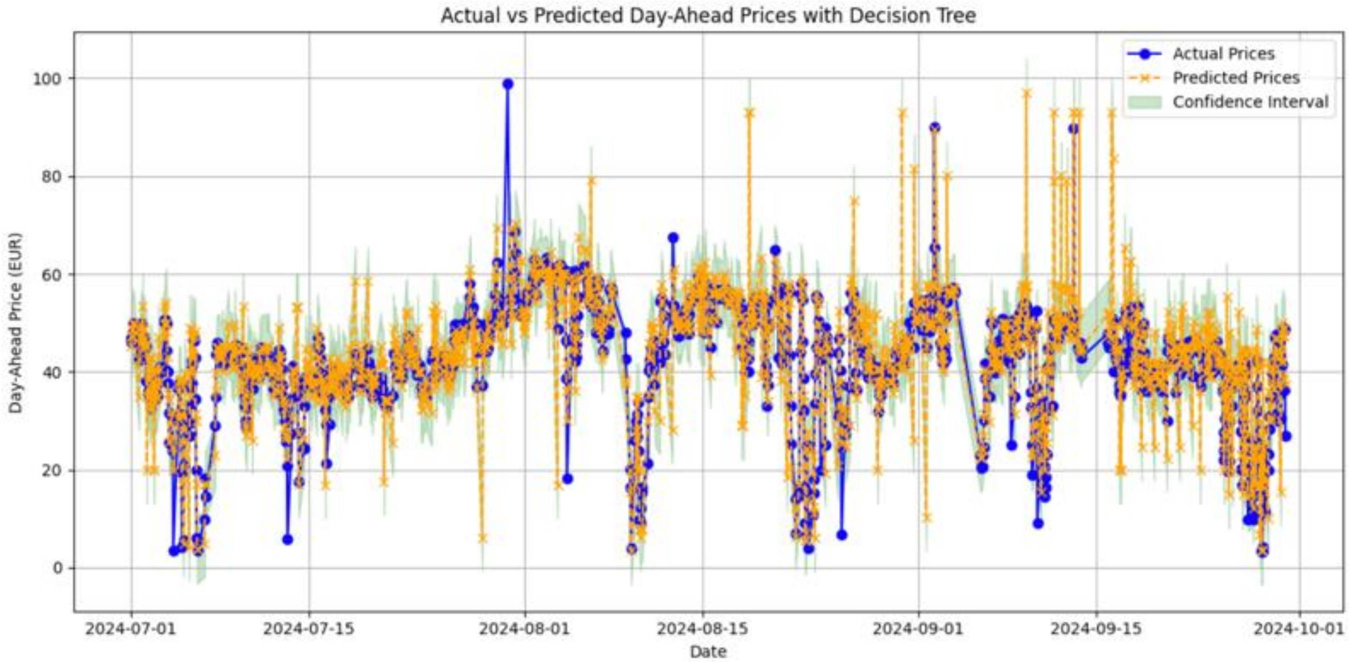

Fig. 12. Actual vs Predicted Day-Ahead Prices with Decision Tree

The predicted price changes in Fig. 12 under the decision tree model are more drastic and fluctuate frequently. During the observed period, the volatility ranges from 0 to 100 Euro. The fluctuation in July was relatively smooth and consistent with the actual price, proving that the market had good stability during this period. Price volatility varied markedly from August to September and there were periods of significant overestimation of prices.

```
Top 10 Feature Importances:
                                 Feature  Importance
0                  Prev_Day_DA_prices_NO_2    0.902195
49  Generation_forecast_NO_2_pct_change    0.031522
19              Generation_forecast_NO_1    0.005301
29                  Net_Flow_NO_2_to_NO_5    0.004148
39  Generation_forecast_NO_2_roll_std_3    0.003926
20              Generation_forecast_NO_5    0.003057
11                         Wind Onshore_DK    0.002826
25        Aggregate_Water_Reservoirs_NO_2    0.002553
35        Generation_forecast_NO_2_lag_30    0.002490
48          Load_forecast_NO_2_pct_change    0.002214
```

Fig. 13. Feature importance based on Decision Tree

The feature importance under the decision tree model in Fig. 13 also shows an obvious hierarchical structure. The score indicates the relative importance of each feature in the overall prediction framework. The decision tree model is highly dependent on “Prev_Day_DA_prices_NO_2”. The importance score of 0.902 indicates that it accounts for 90.2% of the total predictive power of the model. The extreme price increases in the model are probably related to it. When historical prices rise, the decision tree model amplifies the impact of such changes, ultimately leading to aggressive price forecasts.

4.1.3 Random Forest Model

Theory & mathematical foundation

The Random Forest model is an ensemble learning technique that combines the predictions of multiple decision trees to improve predictive accuracy. It works by creating a collection of decision trees during training, where each tree is built using a random subset of the data and a random subset of the features. The final prediction is made by averaging (this can be a weighted average depending on the decision tree) the predictions from all the individual trees, which helps to reduce overfitting and improve generalisation.

Implementation specifics

The random forest model is implemented using the RandomForestRegressor of scikit-learn. It trains multiple trees using bagging and averages the predictions across all trees.

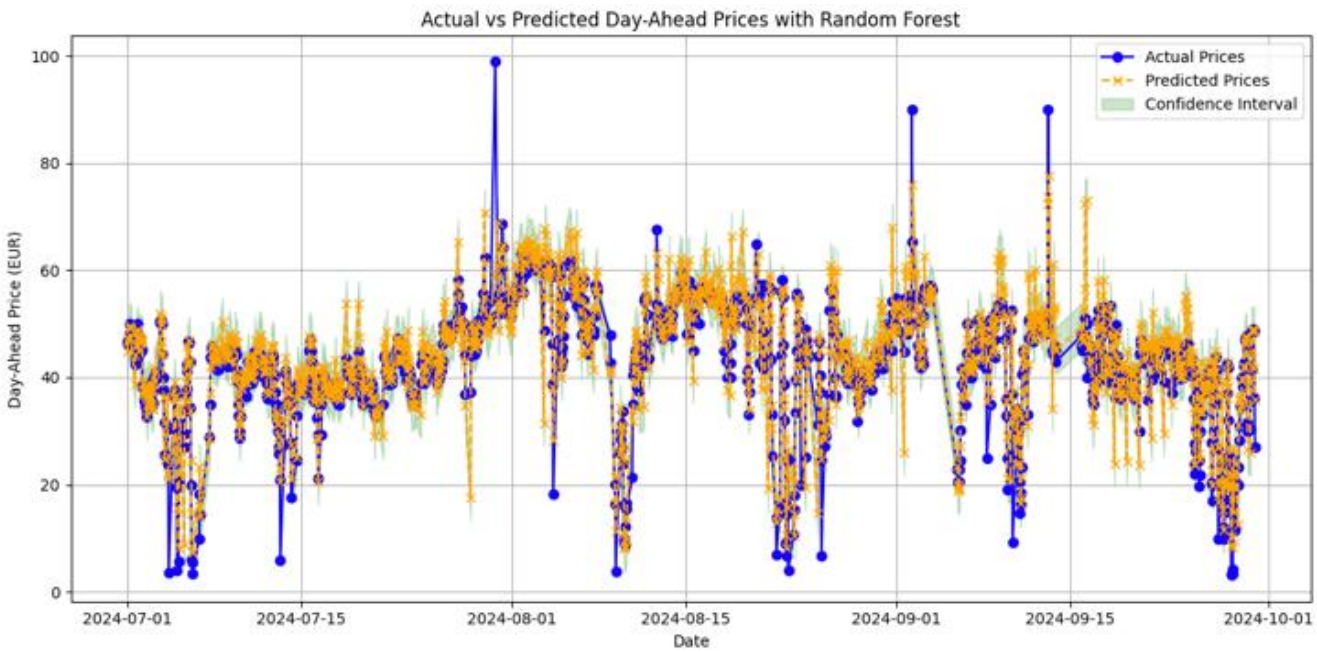

Fig. 14. Actual vs Predicted Day-Ahead Prices with Random Forest

During the observation period in Fig. 14, the predicted price fluctuates from 0 to 80 euro. Except for extreme events, the predicted price volatility under the random forest during the market stability period is basically consistent with the actual price. Notably, random forests also capture the impact of rising prices. But compared to decision trees, random forests are more cautious in their price predictions, predicting peaks that are lower than the actual extremes. Similarly, in the case of declines in actual prices, while the random forest captures these trends but gives a more conservative prediction of prices as well.

Top 10 Feature Importances:

	Feature	Importance
0	Prev_Day_DA_prices_NO_2	0.904712
49	Generation_forecast_NO_2_pct_change	0.030820
19	Generation_forecast_NO_1	0.003492
20	Generation_forecast_NO_5	0.003218
39	Generation_forecast_NO_2_roll_std_3	0.003021
25	Aggregate_Water_Reservoirs_NO_2	0.002759
48	Load_forecast_NO_2_pct_change	0.002559
29	Net_Flow_NO_2_to_NO_5	0.002272
37	Load_forecast_NO_2_roll_std_3	0.001978
18	Generation_forecast_NO_2	0.001898

Fig. 15. Feature importance based on Random Forest

From Fig. 15, the hierarchical structure of feature importance under the random forest model is consistent with that of the decision tree. The percentage of “Prev_Day_DA_prices_NO_2” is 90.4%, which is the main influencing factor. It is worth noting that compared with the decision tree, the random forest enhances the importance of "Generation_forecast_NO_5", proving that it may identify a potential relationship between features.

4.1.4 XGBoost Model

Theory & mathematical foundation

EXtreme Gradient Boosting (XGBoost) is an advanced extension of gradient boosting, where an ensemble of decision trees are built sequentially. The key goal of an XGBoost model is to improve prediction by learning from previous models and building on these prior models with each new tree designed to address the residual errors from earlier iterations. XGBoost is highly efficient and is a popular choice among other machine learning algorithms.

Implementation specifics

The XGBoost model is implemented using XGBRegressor from the xgboost library. It builds multiple decision trees sequentially to minimise prediction errors and averages their outputs.

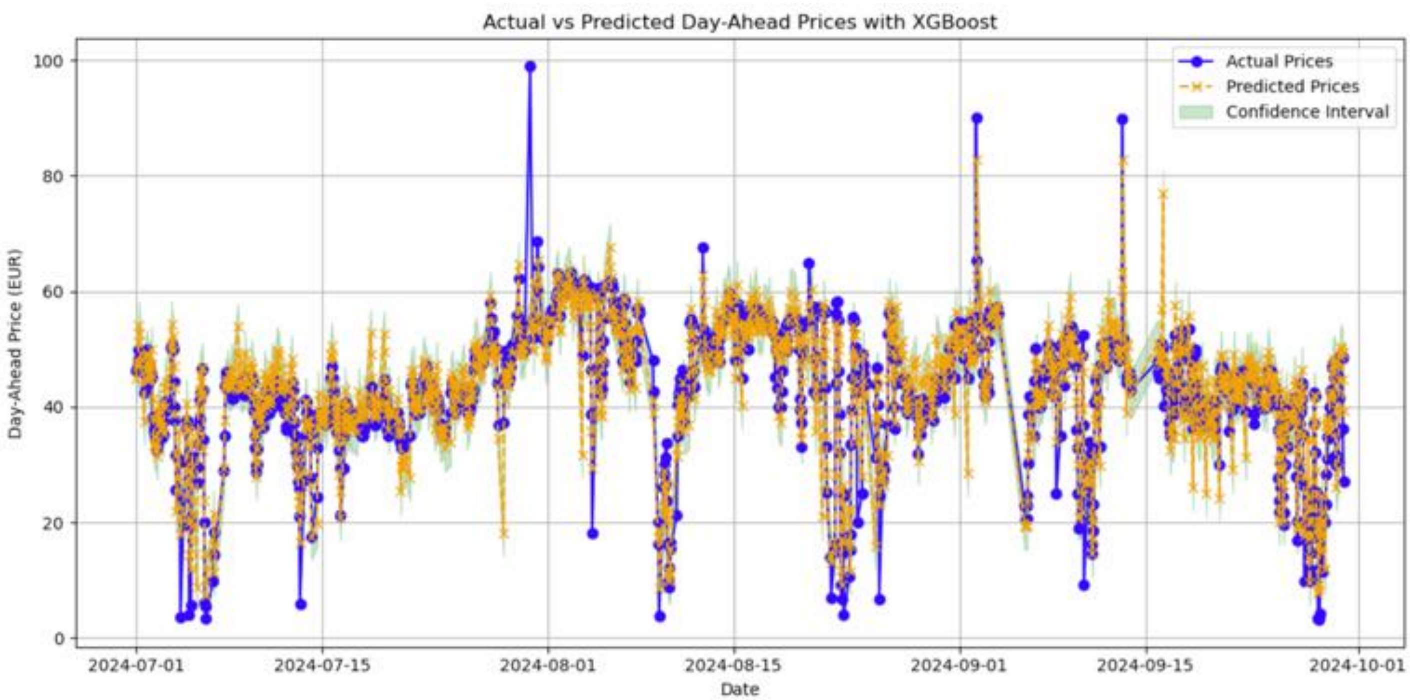

Fig. 16. Actual vs Predicted Day-Ahead Prices with XGBoost

Regarding the XGBoost model in Fig. 16 the predicted price fluctuates from 0 to just over 80 euro. The model seems to predict the price volatility quite accurately for Actual Prices that don't seem to have drastic fluctuations. This model is slightly more conservative which can be seen between August to September.

```
Top 10 Feature Importances:
                                  Feature  Importance
0                     Prev_Day_DA_prices_NO_2    0.549535
18                   Generation_forecast_NO_2    0.133246
40    Generation_forecast_NO_2_roll_mean_3    0.077096
55                                      Month    0.025530
38             Load_forecast_NO_2_roll_mean_3    0.024912
25           Aggregate_Water_Reservoirs_NO_2    0.018266
3                          Load_forecast_NO_5    0.018211
51      Generation_forecast_NO_2_pct_change    0.016966
46            Load_forecast_NO_2_roll_mean_30    0.016671
2                          Load_forecast_NO_1    0.012563
```

Fig. 17. Feature importance based on XGBoost

Interestingly, from Fig. 17 the dependence on "Prev_Day_DA_prices_NO_2" is significantly lower in the XGBoost model compared to the Decision Tree and Random Forest models. While it remains the most important feature at 55.0%, XGBoost allows other features to contribute meaningfully to predictions. Features like **Generation_forecast_NO_2** (13.3%), **Generation_forecast_NO_2_roll_mean_3** (7.7%), and **Month** (2.6%) play key roles.

4.1.5 Support Vector Machine

Theory & mathematical foundation

Support Vector Machine (SVM) is a supervised learning algorithm, most used for classification and regression tasks. SVM works by finding a hyperplane in high-dimensional space, maximizing the separation between the data into distance categories. During regression problems, SVM identifies a function that fits the data within an initialised margin of tolerance with a focus on minimising the errors between the predicted and actual values.

Implementation specifics

The SVM regression model is implemented using SVR from scikit-learn. It fits a hyperplane in a high-dimensional space.

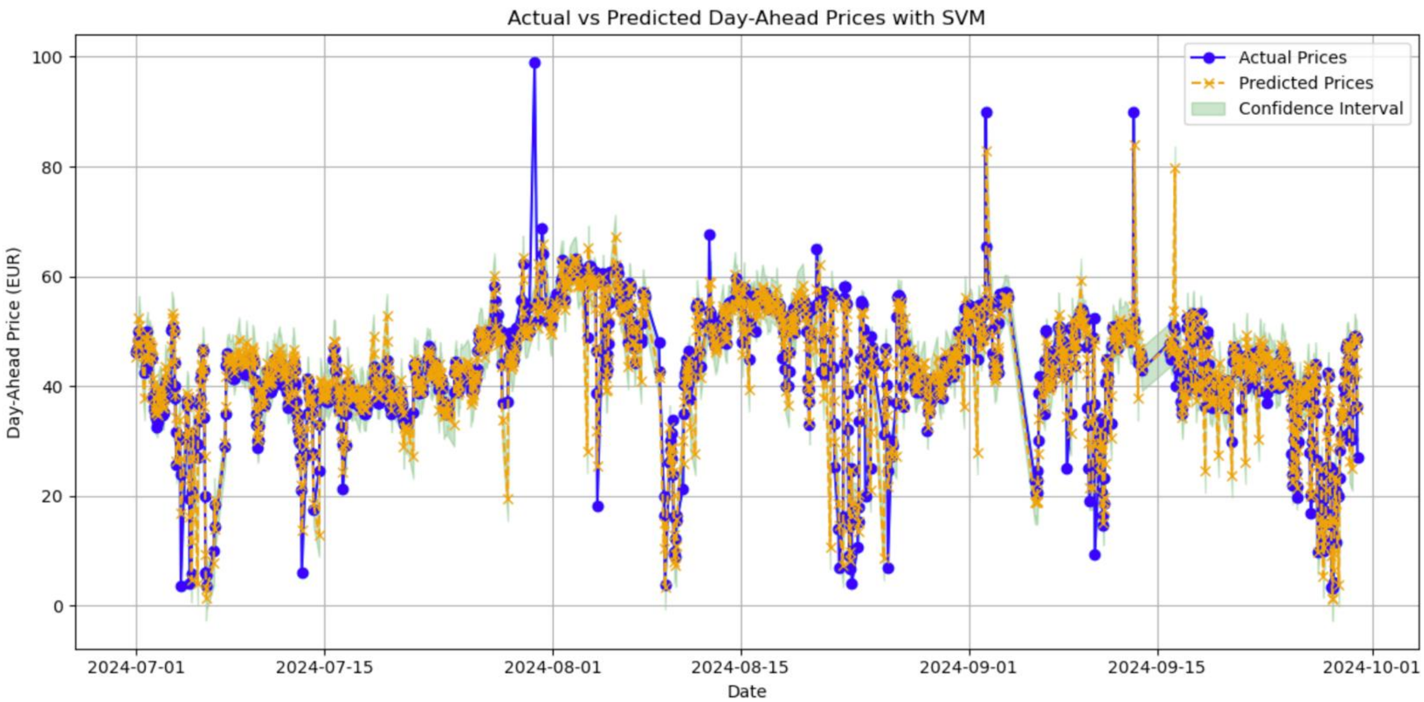

Fig. 18. Actual vs Predicted Day-Ahead Prices with SVM

The SVM model's predicted prices fluctuate between approximately 0 and just over 70 euros in Fig. 18. It seems to capture most of the actual prices quite well. However it struggles to capture larger spikes at times and other times the model can predict false spikes. This indicates that the SVM may not be fully capturing the dynamic nature of price fluctuations, particularly the sharp increases, which could be crucial for accurate forecasting in markets with volatile price movements.

Top Feature Importances:

	Feature	Importance
0	Prev_Day_DA_prices_NO_2	0.466695
51	Load_Generation_ratio	0.035778
52	Load_Generation_diff	0.027246
50	Generation_forecast_NO_2_pct_change	0.020922
9	Wind Offshore_DK	0.018421
8	Solar_DK	0.016276
18	Generation_forecast_NO_1	0.016186
17	Generation_forecast_NO_2	0.016057
10	Wind Onshore_DK	0.013467
20	Generation_forecast_DK	0.012046

Fig. 19. Feature importance based on SVM

From Fig. 19 the SVM model identifies the most important feature as "Prev_Day_DA_prices_NO_2" at 46%, the importance of the other features drop drastically after this.

4.2 Performance Metrics

Performance metrics are used to evaluate the accuracy of the models by measuring the errors between predicted and actual values, helping to assess model quality.

4.2.1 Theory & mathematical foundation

There are three main model evaluation metrics that were used to evaluate the performances of the models. These included Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared (R^2). A brief outline of how these performance metrics works are shown below:

Mean Squared Error

This statistical technique evaluates the difference between the target values and the predicted values. The equation is given in Equation 1.

Equation 1

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

Where is the target value the predicted and the number of data points. The square is taken for several reasons. One of which being to ensure the negative MSE values don't cancel the positive. The squared term also ensures that large discrepancies in the predicted versus target are penalised and pointed out clearly. A lower MSE value indicates a better prediction as it shows that the predicted values and closer to the target values. This metric is widely used due to it's comprehensive nature.

Mean Absolute Error

This metric calculates the sum of the average differences between the predicted and actual values. The equation is given in Equation 2.

Equation 2

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

This equation is very similar to the MSE however the residuals are not squared, instead the absolute value is taken. Unlike MSE the MAE is resilient to outliers and by not penalising large differences between the predicted and target values. The two metrics can be used together to give insights into the performance of the models. If MAE is much smaller than MSE it indicates that most errors are small but there are a few larger ones, and possibly some outliers which shoot the value of the MSE up.

R-Squared and Adjusted R-Squared

This measure represents the proportion of the variance in y is captured by the model as opposed the variation around the average value of y. the equation is given as in Equation 3.

Equation 3

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)}{\sum_{i=1}^{n}(y_i - \bar{y}_i)}$$

this value lies between 0 and 1, with 1 indicating the perfect fit to the data (explaining all the variance in the dependent variable). R^2 is not usually used as a stand-alone metric and is compared between different models and used alongside different metrics to better obtain a well-rounded view of model performances. There is a risk of overfitting when it comes to this metric as it only increases or keeps a constant with the introduction of more predictors. The adjusted R^2 is an extension of R^2 that penalises predictors that are not meaningful to the data.

4.2.2 Result of Metrics

Table 1. Evaluation Metrics across the different Models

Model	**MSE**	**MAE**	**R^2**	**Adjusted R^2**
Linear Regression	17.88	2.59	0.85	0.84
Decision Tree	24.25	3.06	0.8	0.79
Random Forest	16.96	2.33	0.86	0.85
XGBoost	16.33	2.62	0.86	0.85
SVM	15.19	2.18	0.87	0.86

5. Results Analysis and Model Selection

5.1 Predicted vs. Actual plots

Regarding the ability of different models in price prediction above, linear regression can capture the price trend in the steady period, but it is more likely to have a large bias when the price fluctuates. Decision trees have the largest bias throughout the period. The Random Forest is more conservative in it's predictions, especially for extreme prices as well as peak tariffs. In contrast the SVM and the XGBoost perform optimally. However, all models have deviations in the low-price prediction.

5.2 Best Features

When comparing the best features from the theory and results section it is clear that "Prev_Day_DA_prices_NO_2" was ranked, excluding the Linear Regression Model which placed the most importance on "Load_forecast_NO_2_pct_change" followed by "Generation_forecast_NO_2_pct_change". The reason for this could be due to the fact that the Linear Regression model assumes a linear relationship between the features and the target variables. Therefore If a feature is highly correlated with the target variable in a linear way, it will likely be deemed the most important by this model, which is what might be seen by the feature "Load_forecast_NO_2_pct_change".

For both the Decision Tree and the Random Forest the "Prev_Day_DA_prices_NO_2" seemed to dominate the as the most important feature which accounts for 90% of the total for both models. This can be seen as a limitation in the model as relying so heavily on a singular feature may cause instability in the predictions.

While the "Prev_Day_DA_prices_NO_2" was still ranked as the most important feature among the XGBoost and the SVM models, their feature importances were more evenly distributed. This could be due to these models being able to capture more complex patterns that may be missed by the other models.

It is also worth noting that certain temporal characteristics are included in all model predictions. The feature importance under linear regression and XGBoost clearly indicates the impact of "month". Decision trees and random forests also cover time-specific features such as "Generation_forecast_NO_2_roll_std_3".

5.2 Performance Metrics

Based on the evaluation metrics shown in Table. 1, the SVM model emerges as the best performing model with a lowest MAE and MSE values at 15.19 and 2.18 respectively. This model also has the highest R^2 and adjusted R^2 values just creeping above the XGBoost and the Random Forest models, which both come which are just slightly behind in terms of performance. The worst performing model is the decision tree giving a large MSE value of 24.25, this value can be reflected in Fig. 12 where many inaccurate price spikes were In contrast, the Decision Tree model is the poorest performer, with a significantly higher MSE value of 24.25. This result is evident in Fig. 12, where numerous inaccurate price spikes predicted by the model can be observed. The Linear Regression model, although not as effective as most other models, performs reasonably well considering its simplicity and the minimal optimisation applied.

6. Comparative Analysis with Existing Literature

Lago et Al. (2018) presented the results of open-source, state-of-the-art benchmark models across European electricity markets, for the purposes of evaluation of future research. The models used were four variations, each of a deep neural network (DNN) and a Lasso Estimated Autoregressive model (LEAR). For the Nordic countries, the results can be found in Table. 2.

Table 2. The MAE values obtained for the different DNN and LEAR models found from Lago et Al. (2018)

	DNN_1	DNN_2	DNN_3	DNN_4	$LEAR_{56}$	$LEAR_{84}$	$LEAR_{1092}$	$LEAR_{1456}$
MAE	1.797	2.118	**1.712**	1.883	1.964	1.952	1.993	1.99

In comparison, the best performing model in our study is the SVM model with an MAE of 2.18. This shows that the models in our study do not perform better than state-of-the-art DNN and LEAR models. However, it also illustrates that even simple models with relatively few features can come close to state-of-the art performance.

6.1 Limitations and Future Improvements

A key limitation in our study is the use of only simple supply and demand fundamentals from ENTSO-e. Using financial data on carbon prices, gas prices, oil prices, geopolitical and economic uncertainty etc., could improve model performance. For example, Bendiksen and Løining (2024) used 109 variables in their OLS model for NO2 Day-ahead prices, far more than have been considered here.

Future research could gather a more comprehensive dataset and test on the same models as tested here, to compare the performance of these models with a higher-dimensional dataset. Furthermore, we have not tested neural networks in this study. Further research can compare the performance of various neural networks to the results found here.

7. Conclusion

The objective of this project was to predict day-ahead electricity prices in the NO2 bidding zone using data from ENTSO-e. The inquiry went into the primary elements impacting power pricing by preparing data and developing features. These factors included load forecasts, renewable generation, and cross-border electricity flows.

Our analysis showed that machine learning models like the SVM, XGBoost and Random Forest performed better than traditional methods such as Linear Regression and Decision Trees. Particularly, SVM returned high predictive accuracy and was able to capture complex patterns in the data. However, it should be noted that certain challenges did persist. The models faced difficulty in predicting extreme price fluctuations and resulted in weaker performance during unusual price fluctuation.

This experiment presented the use of simple supply and demand data for predicting electricity prices. While these approaches may not be as effective as sophisticated deep learning algorithms, they produce consistent and interpretable results with fewer resources. In the future, additional data such as weather predictions or fuel costs may increase accuracy and limit the highlighted constraints.

In conclusion, this study offers useful insights on electricity price forecasts. It highlights how available data and machine learning approaches might help market players and governments make proper decisions.

References

ACER (2024) ‘Bidding Zone Review’. Agency for the Cooperation of Energy Regulators. Available at: https://www.acer.europa.eu/electricity/market-rules/capacity-allocation-and-congestion-management/bidding-zone-review (Accessed: 9 January 2025)

Bendiksen, V. and Løining, L.O.F. (2024) ‘Analyzing drivers of day-ahead electricity prices in Norway NO2: Is there an impact of cross-zonal connections?’ MS thesis. Inland Norway University.

Borenstein, S., (2001). ‘The trouble with electricity markets (and some solutions)’. Berkeley: University of California Energy Institute.

Bunn, D.W. (2004) ‘Modelling prices in competitive electricity markets’. United Kingdom: Wiley

ENTSO-E (2024a) ‘Water Reservoirs and Hydro Storage Plants’ Available at: https://transparency.entsoe.eu/generation/r2/waterReservoirsAndHydroStoragePlants/show (Accessed: 7 January 2025)

ENTSO-E (2024b) ‘Installed capacity per production type’ Available at: https://transparency.entsoe.eu/generation/r2/installedGenerationCapacityAggregation/show (Accessed: 7 January 2025)

Fabra, N. (2023) ‘Reforming European electricity markets: Lessons from the energy crisis’. *Energy Economics*, 126, 106963. Available at: https://doi.org/10.1016/j.eneco.2023.106963 (Accessed 7 January 2025)

Hirth, L., Mühlenpfordt, J. and Bulkeley, M. (2018) ‘The ENTSO-E Transparency Platform – A review of Europe’s most ambitious electricity data platform’, *Applied Energy*, 225, pp. 1054–1067. Available at: https://doi.org/10.1016/j.apenergy.2018.04.048 (Accessed 6 January 2025)

Høyslo, J.G. and Hjertner, T., (2022) ‘Analysis of the Norwegian power market with prediction of day-ahead spot prices by means of Linear Regression and Machine Learning’ (Bachelor's thesis, NTNU).

Jędrzejewski, A., Lago, J., Marcjasz, G. and Weron, R. (2022) ‘Electricity price forecasting: The dawn of machine learning’, *IEEE Power and Energy Magazine*, 20(3), pp. 24–31. Available at: https://doi.org/10.1109/MPE.2022.3150809 (Accessed 4 January 2025)

Lago, J., Marcjasz, G., De Schutter, B. and Weron, R. (2021) ‘Forecasting Day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark’, *Applied Energy*, 293, 116983. Available at: https://doi.org/10.1016/j.apenergy.2021.116983 (Accessed 4 January 2025)

Nord Pool (2024) ‘Day-ahead Prices’. Available at: https://data.nordpoolgroup.com/auction/day-ahead/prices?deliveryDate=latest&currency=NOK&aggregation=YearlyAggregate&deliveryAreas=NO1,NO2,NO3,NO4,NO5 (Accessed 7 January 2025)

Norwegian Ministry of Petroleum and Energy (2008) ‘Facts 2008: Energy and Water Resources in Norway - Chapter 7’. Available at: https://www.regjeringen.no/globalassets/upload/oed/pdf_filer/faktaheftet/evfakta08/evfacts08_kap07_eng.pdf (Accessed 3 January 2025)

Rathke, L.M. (2015) ‘The Effects of Electricity Market Liberalisation in the European Union’. Bachelor's thesis, University of Twente

Richard Milne (2024) ‘Norway campaigns to cut energy links to Europe as power prices soar’. *Financial times*. Available at: https://www.ft.com/content/f0b621a1-54f2-49fc-acc1-a660e9131740 (Accessed 8 January 2025)

Viken, O.P. (2024) ‘Analysing Norwegian Electricity Prices During the Energy Crisis: A Data-Driven Forecasting Approach with Explainable AI’ MS thesis. Norwegian University of Life Sciences.

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), pp. 1030–1081. Available at: https://doi.org/10.1016/j.ijforecast.2014.08.008 (Accessed 3 January 2025)

Wikipedia Commons (2022) ‘Electricity Price Ara Map’. Available at: https://commons.wikimedia.org/wiki/File:Hou710_ElectricityPriceArea.svg (Accessed: 8 January 2025)

Zakeri, B., Staffell, I., Dodds, P.E., Grubb, M., Ekins, P., Jääskeläinen, J., Cross, S., Helin, K., and Castagneto Gissey, G. (2023) ‘The role of natural gas in setting electricity prices in Europe’. Energy Reports, 10, pp. 2778–2792. Available at: https://doi.org/10.1016/j.egyr.2023.09.069 (Accessed 7 January 2025)

Contribution table					
UCL candidate code	**Role in project**	**Major contributions**	**Challenges faced and overcome**	**Hours contributed**	**Additional notes**
LSHM1	Researcher	Responsible for introduction and context, literature review and data retrieval through ENTSO-e API service	Main challenge was to understand the field of electricity price forecasting and identify suitable data for our project	15	
PKHJ8	Research support	Supported research and produced charts	Understanding electricity price forecasting and navigating the API service	15	
MHCR1	Data Analyst	Responsible for EDA process, retrieving and cleaning data from the API, selecting features, and preparing datasets for analysis.	Faced challenges with varying data resolutions, missing values, and feature alignment, overcome through careful preprocessing and flexible methodologies	15	
LDYY7	Model Trainer	Responsible for testing the data with different models and hyper tuning the models to achieve better performances ultimately comparing the models	Figuring out how to split different parameters to test into smaller groups so the GridsearchCV was less computationally expensive	15	
PDXT9	Model Analyst	Responsible for Model Diagnostics and training, analyzing the model by illustrating the graph, comparing results of different models	Facing challenges with model illustrating and comparing, figuring out by organizing visual comparisons through charts	15	