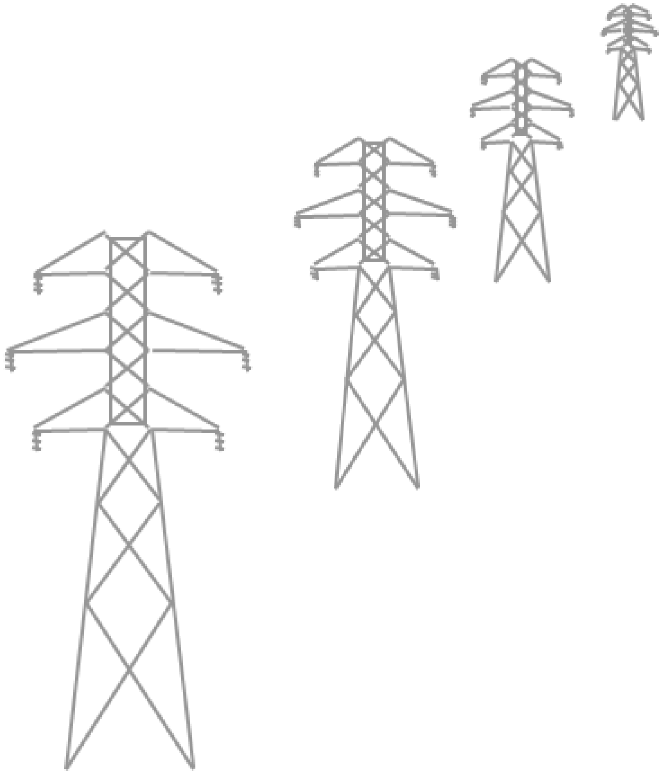


Subject:

Day-ahead Electricity Price Forecasting (Case study: Germany)

Presented By:

Naser Rostamnia

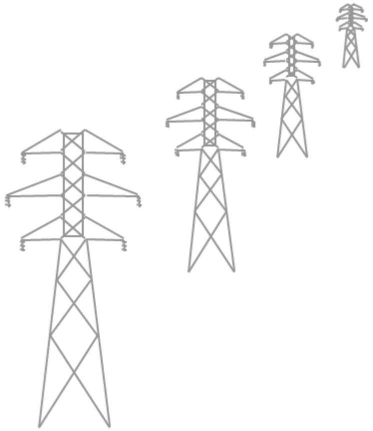
Qualifications:

MSc. Data Science, AI and Digital Business

MSc. Energy Systems Planning

BSc. Applied Mathemtics

July 2024



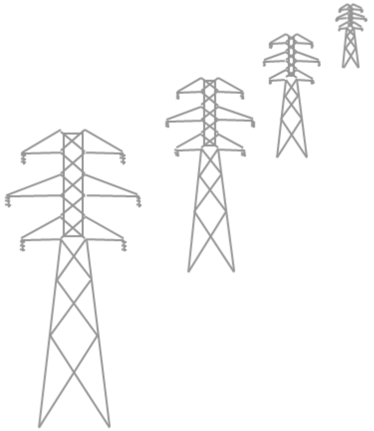
Introduction:

Forecasting electricity prices is crucial for energy companies and the effective management of smart grids. Due to the fluctuating nature of electricity demand, prices can exhibit sudden, short-lived, and often unpredictable spikes. Accurate predictions of both price levels and volatility are essential for decision-makers, enabling them to fine-tune bidding strategies and optimize production or consumption schedules.

Objective: Analyzing the day-ahead electricity prices to identify key trends and influencing factors.

Scope:

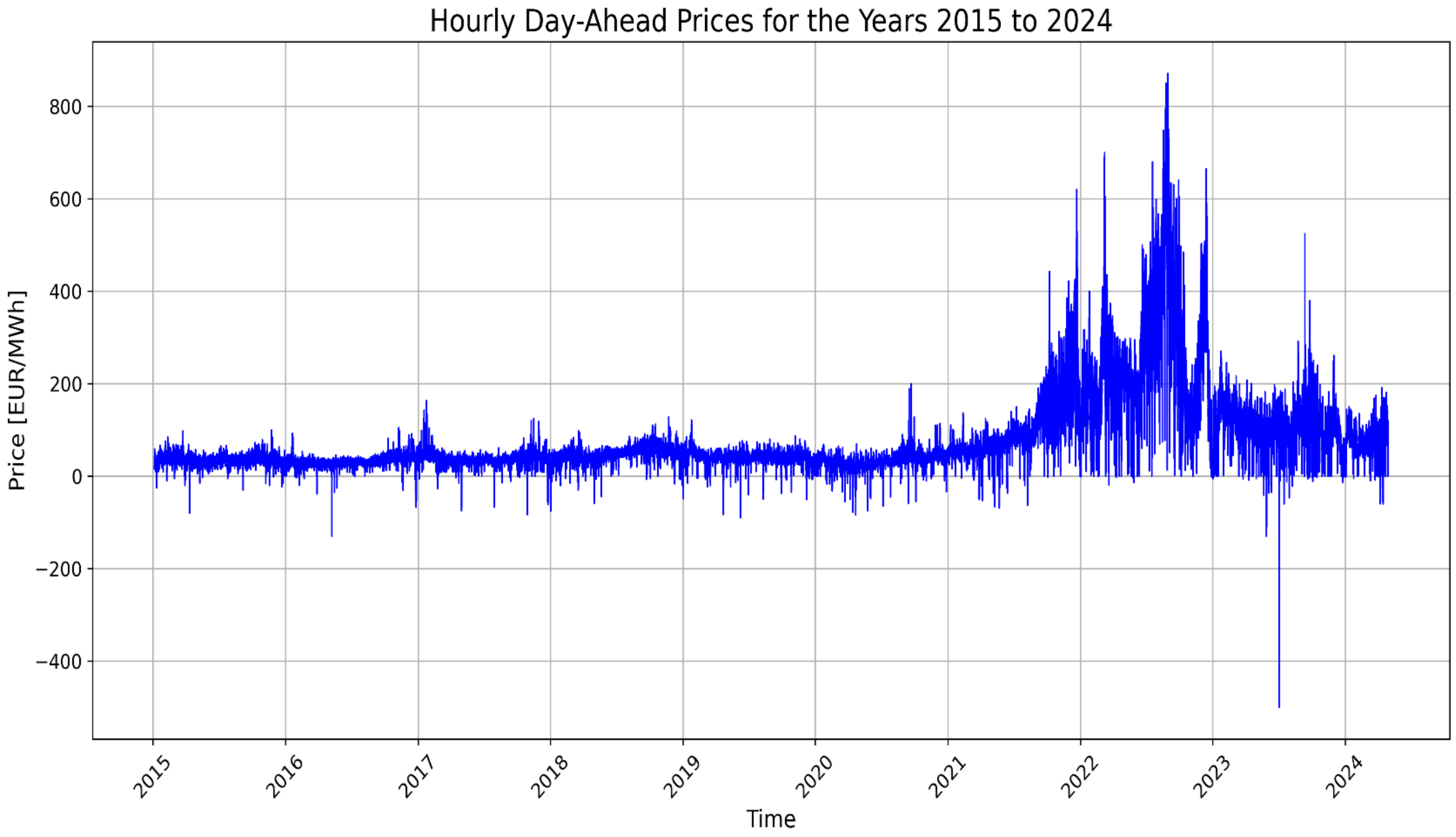
- **Geographical Coverage**: Germany-Luxembourg(+Austria until Oct.2018),
- **Time Period:** The dataset spans from January 2015 to April 2024
- **Total number of Hourly Observations:** 81695 hours

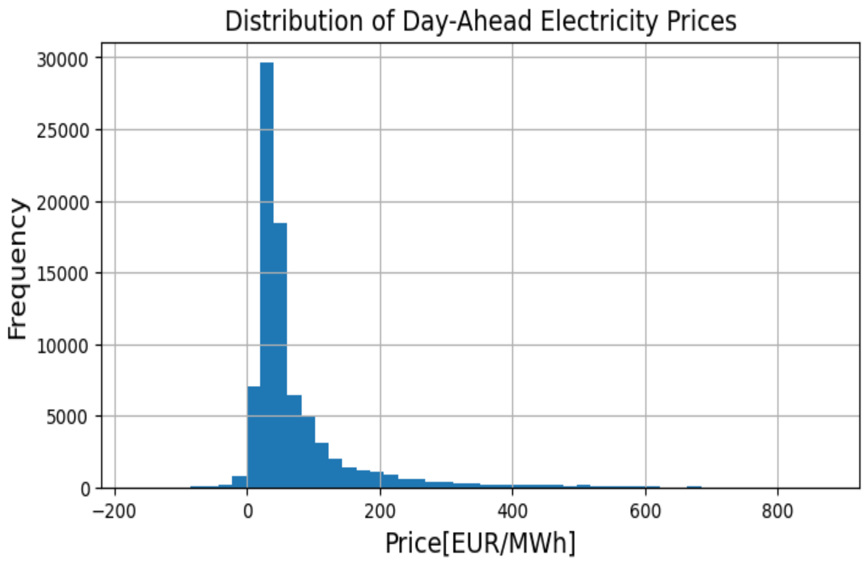


Feature summary with key statistics

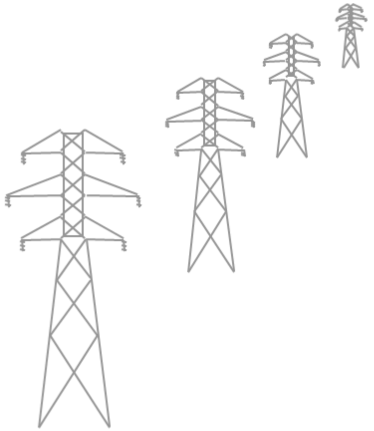
Feature Name	**Null Values**	**Mean**	**Median**	**Min**	**Max**	**Std Dev**
DateTime	0	-	-	-	-	-
Price[EUR/MWh]	73	70.46	42.00	-500.00	871.00	84.30
Load_DA[GW]	2	54.82	54.71	28.82	77.59	9.38
Load_Actual[GW]	20	56.31	56.02	30.91	81.08	9.98
Solar_DA[GW]	0	4.98	0.18	0.00	41.17	7.75
Solar_Actual[GW]	20	4.97	0.14	0.00	40.67	7.77
WindOnshore_DA[GW]	0	2.31	1.99	0.00	6.77	1.78
WindOnshore_Actual[GW]	20	2.34	2.02	0.00	7.63	1.78
WindOffshore_DA[GW]	0	10.54	7.85	0.07	46.62	8.69
WindOffshore_Actual[GW]	20	10.62	7.90	0.07	48.02	8.80
Temperature_DA[C]	54,551	10.94	10.40	-11.75	35.62	7.46
Temperature_Actual[C]	0	10.94	10.40	-11.75	35.62	7.46
Coal_fM	79,264	58.87	55.72	17.80	117.49	18.31
Gas_fD	79,263	34.55	19.98	3.10	330.00	29.97
Gas_fM	79,263	35.07	19.68	3.54	310.50	41.01
Gas_fQ	79,263	36.12	19.16	4.75	331.39	42.22
Gas_fY	79,263	33.68	19.19	11.83	305.00	35.40
Oil_fM	79,264	58.87	55.72	17.80	117.49	18.31
EUA_fM	79,264	34.55	24.18	3.93	97.59	29.97
EUR_USD	79,263	1.12	1.12	0.96	1.25	0.05

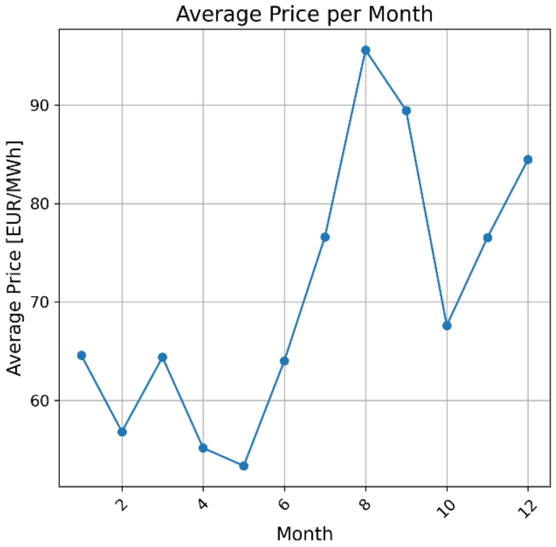
Data Visualization

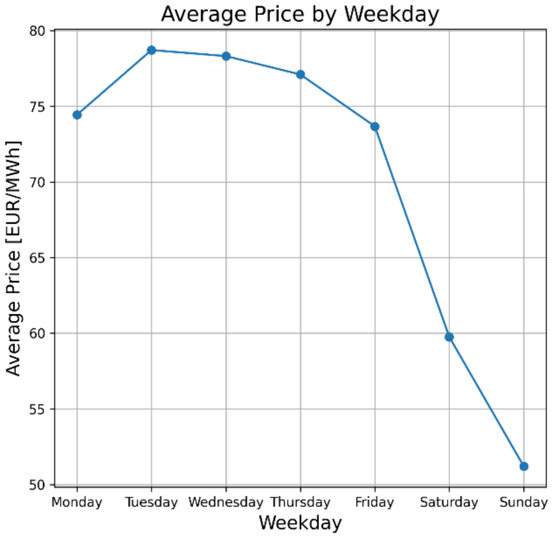


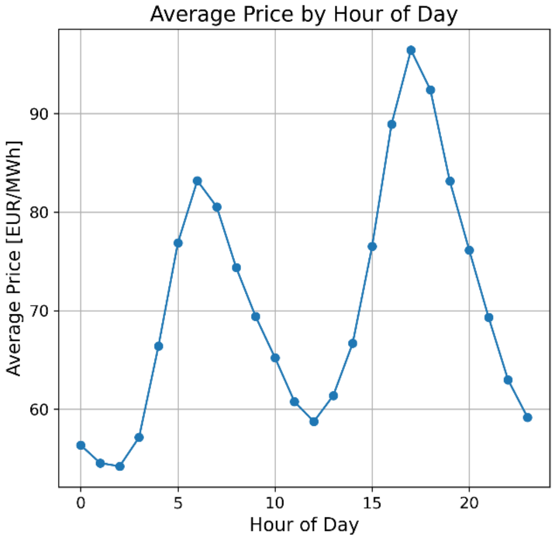


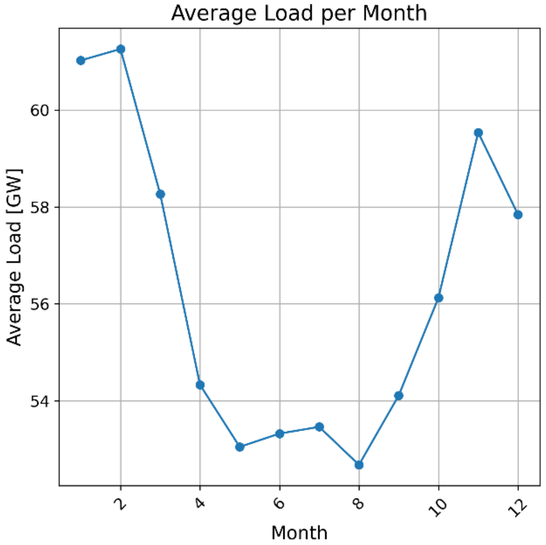
Average Load and Price: Hourly, Weekly and Monthly

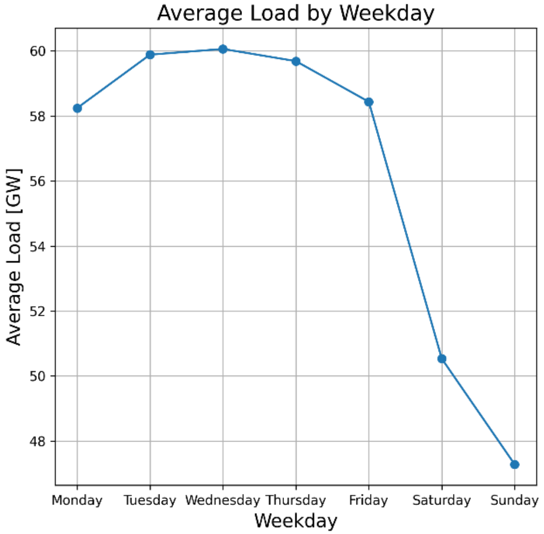


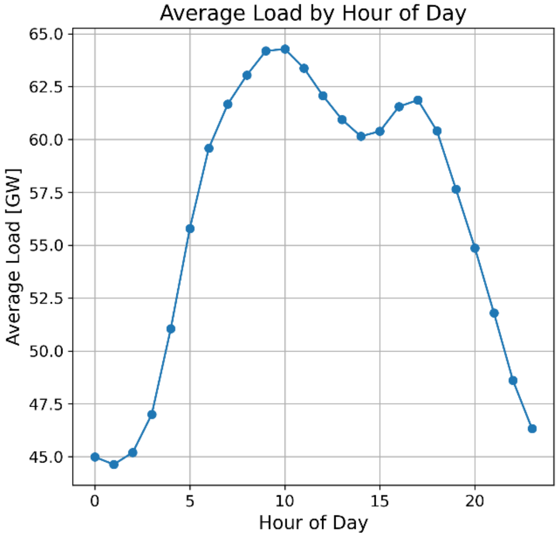




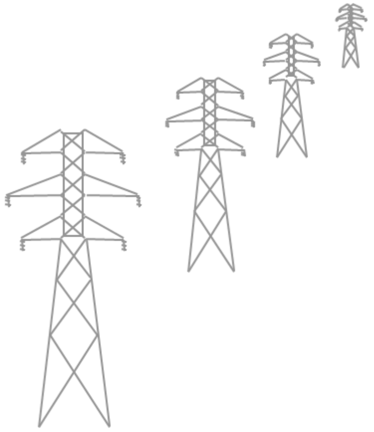


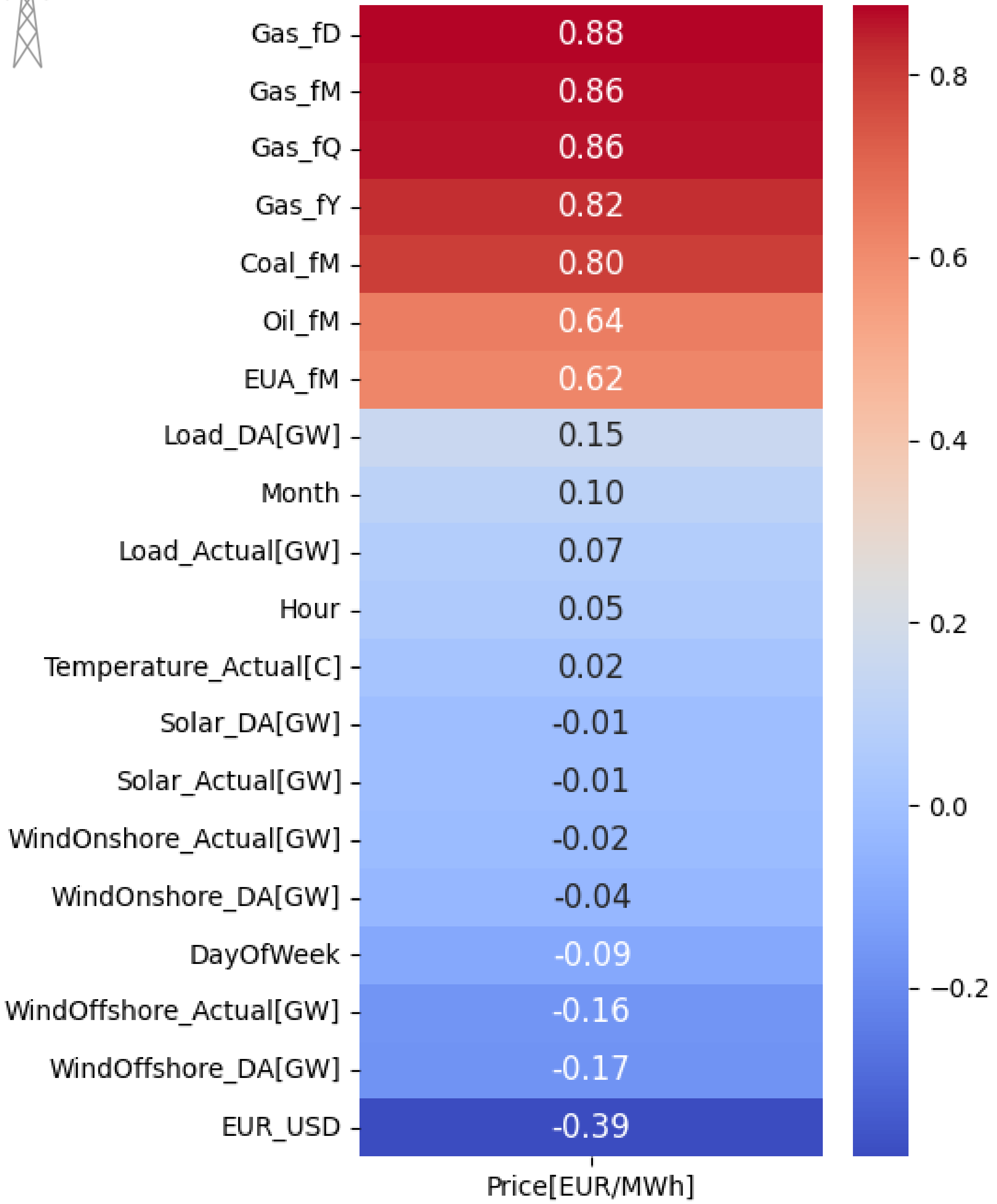


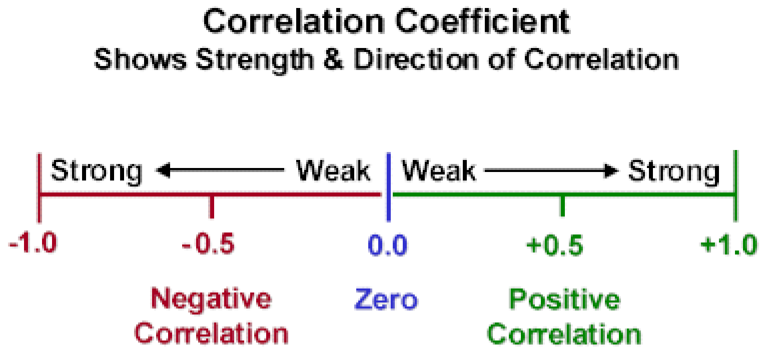




Correlation Coefficient







number	Feature Used in this study
1	Historical Price (d-1)
2	Load_DA[GW]
3	WindOffshore_DA[GW]
4	WindOffshore_Actual[GW]
5	Coal_fM
6-9	Gas_fD, Gas_fM, Gas_fQ, Gas_fY
10	Oil_fM
11	EUA_fM
12	EUR_USD

Data Preprocessing steps:

- **Handling missing values**
  - ✓ Remove Last 72 Rows: Exclude the last 72 rows (with missing price values) as they do not interfere with the time serries order
  - ✓ Drop 'Temperature_DA[C]' due to limited data.
  - ✓ Forward Fill Strategy: Apply forward filling to handle missing values in remaining features

- **Mitigating the effect of extreme values**
  - ✓ Replace values less than -200 EUR/MWh with the price recorded at 10:00, which was -167.96 EUR/MWh

- **Adding extra features to the data set:**
  - ✓ Hour, Day of Week and Month.

- **Calculating Correlation Coefficient as a guide for feature selection:**

- **Data standardization:**
  - ✓ Rescale data to mean 0, std dev 1.

- **Using two different strategies for splitting data into Training, validation and Test subsets:**
  - ✓ Train_test_split from sklearn laberatory
  - ✓ Sliding window method (creating windows with size 24, to predict the price for the next 24 hours.)
  - ✓ Training: 80%, Validation: 10% and Test :10%

Model Building, Training, and Evaluation

Two different models are used for day-ahead price prediction:

i. Recurrent Neural Network (RNN):

- ✓ **Model Architecture: LSTM layers** for temporal dependencies, **Dense layers** for predictions.
- ✓ **Training Parameters:**

  Optimizer: Adam,

  Learning Rate: {0.01, 0.001},

  Epochs: 500,

  Batch Size: 256.
- ✓ **Performance Metric:** Mean Absolute Error(MAE)

ii. RandomForestQuantileRegressor:

- ✓ **Quantiles:** 10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th}.
- ✓ **Performance Metric:** Mean Absolute Error(MAE)
- ✓ **Data splitting strategies:**

  Sliding window method (creating windows with size 24, to predict the price for the next 24 hours.)

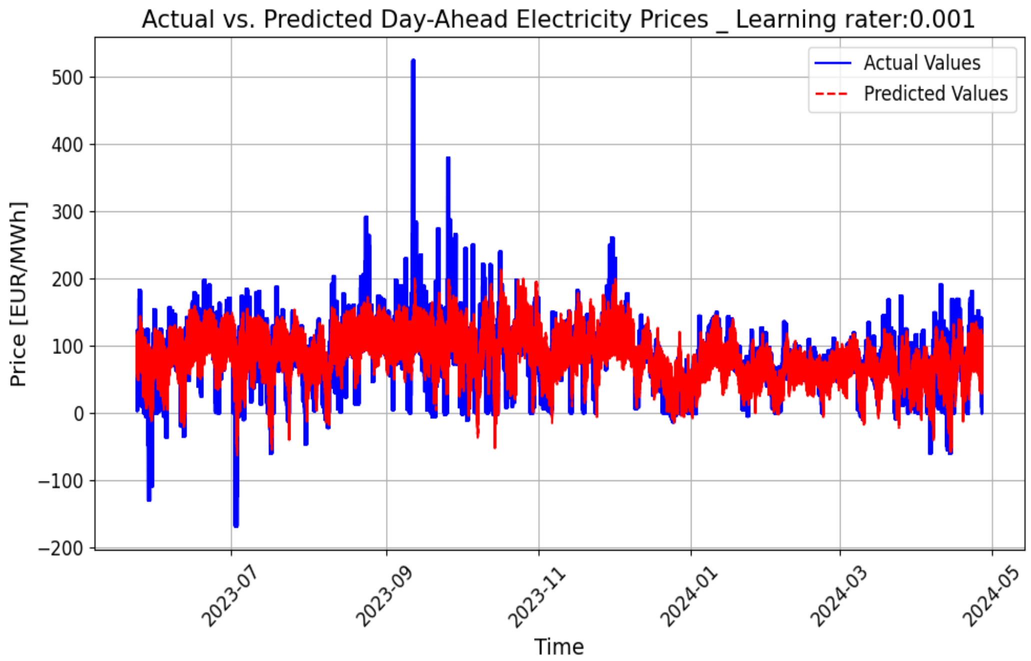
  Train_test_split method from sklearn liberatory.

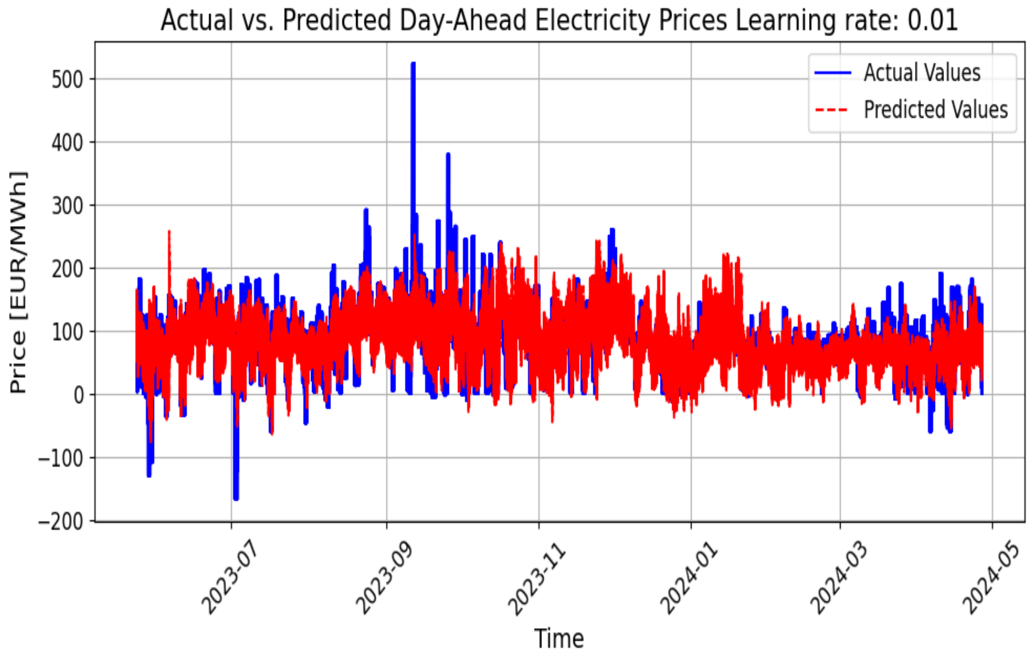
RNN Model, Performance and visualization

```
Layer (type)                Output Shape              Param #
=================================================================
lstm_2 (LSTM)               (None, 256)               275456

repeat_vector_1 (RepeatVec  (None, 24, 256)           0
tor)

lstm_3 (LSTM)               (None, 24, 256)           525312

time_distributed_1 (TimeDi  (None, 24, 1)             257
stributed)

=================================================================
Total params: 801025 (3.06 MB)
Trainable params: 801025 (3.06 MB)
Non-trainable params: 0 (0.00 Byte)
```

MAE Performance Evaluation of RNN

Learning rage	MAE
0.01	27.9
0.001	19.8

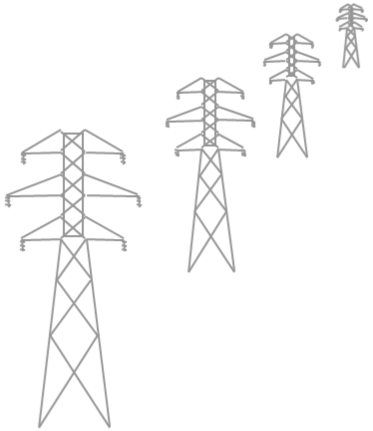




RandomForestQuantileRegressor Model, Performance and visualization

Performance Comparison of RandomForestQuantileRegressor		
Quantile	Sliding Window Strategy MAE	Train-Test Split Strategy MAE
10th Percentile	39.1	15.9
25th Percentile	28.3	10.7
50th Percentile	24.7	8.7
75th Percentile	31	11
90th Percentile	41	15.9

Strategy	Actual Price	Predicted 10^{th} Percentile	Predicted 25^{th} Percentile	Predicted 50^{th} Percentile	Predicted 75^{th} Percentile	Predicted 90^{th} Percentile
Sliding window	77.1	71.6	74.8	79.9	84.1	90.0
	69.6	69.5	73.7	78.1	83.3	87.0
	70.0	69.1	73.9	80.5	84.6	94.9
	78.6	69.2	74.3	83.2	92.3	110.1
	101.1	70.3	78.0	94.1	106.9	122.3
Train Test split	26.4	30.4	30.4	39.7	44.8	46.7
	29.7	26.5	28.0	31.3	32.9	36.3
	383.9	367.2	374.7	382.9	397.9	398.9
	53.9	48.1	50.1	54.3	60.4	60.4
	11.9	10.5	14.9	14.9	24.5	31.6



Feature Engineering and Model Optimization

- Electricity prices are influenced by a range of factors, both directly and indirectly. Some of these factors are included in the dataset, while others can be derived from the existing data. For instance, incorporating historical prices (e.g., prices from 1, 2, 3, or 7 days prior) and lagged values of other features could enhance prediction accuracy.

- Feature engineering is crucial for improving prediction results, as it involves creating relevant and meaningful features from raw data. Unfortunately, due to time constraints, I wasn’t able to explore various scenarios and combinations of these factors as thoroughly as I would have liked. However, I recognize the importance of this process.

- Additionally, optimal results are often achieved by implementing several models and tuning their hyperparameters. By experimenting with different models and selecting the one that performs best, one can significantly improve the prediction accuracy.

Reference

i. Chou, J. S., & Ngo, N. T. (2016). Time series analytics using sliding window metaheuristic optimization-based machine learning system for identifying building energy consumption patterns. Applied energy, 177, 751-770.

ii. Geron, A. (2019), Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems, O'Reilly Media.

iii. Juban, R., Ohlsson, H., Maasoumy, M., Poirier, L., & Kolter, J. Z. (2016). A multiple quantile regression approach to the wind, solar, and price tracks of GEFCom2014. International Journal of Forecasting, 32(3), 1094-1102.

iv. Rostamnia, N., & Rashid, T. A. (2019). Investigating the effect of competitiveness power in estimating the average weighted price in electricity market. The Electricity Journal, 32(8), 106628.

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

Thank you