Hochschule Flensburg

University of Applied Science

Wind Energy Technology Institude

Research on wind Energy

Power Prediction and Energy Arbitrage Modelling for Wind Farms Using Machine Learning

Market Integration and Commercial Optimization, WiSe 2025/26

Master's degree in Wind Engineering, Flensburg, Germany

Submitted by

Karan Soni   (760153)
Madhvesh Gorakhiya (730251)

Supervising Professors:

Prof. Dr. Andreas Heinen

Date of submission:

26.01.2025

1. Abstract

Objective, methodology, and key findings

Highlight ML approach and arbitrage results

2. Introduction

Wind energy is one of the fastest-growing renewable industries in Germany. It plays a critical role in reducing carbon emissions and achieving sustainability targets. [1] However, wind power generation is variable because wind speeds fluctuate continuously, making it difficult to predict power output accurately. This variability creates significant challenges for electricity markets, where prices also fluctuate strongly. Wind production is high during low-price periods, selling electricity immediately can reduce profits for wind-farm operators. (Add reference at here) Storing energy for later sale may increase revenue, but deciding when to store or sell is complex.

Machine Learning (ML) is good suited for improve forecasting and operational decisions. [2] ML models can capture nonlinear patterns in wind speed and power output, enabling more accurate predictions compared to traditional statistical methods. Forecasting is essential not only for grid stability but also for implementing energy arbitrage strategies, which allow operators to maximize profits during price fluctuations. [3] However, uncertainty in electricity prices and wind power output makes revenue prediction challenging. Selling at the wrong time or storing energy unnecessarily can lead to financial losses.

This report investigates two critical research questions:

1. Can machine-learning models accurately predict wind-farm power output and identify the optimal hours to STORE or SELL energy?
2. If low-price energy is stored as hydrogen, can the hybrid wind-hydrogen system remain economically profitable?

The primary objectives of this study are to develop and evaluate ML-based models for wind power forecasting and to assess the economic evaluation of hydrogen-based energy storage in arbitrage scenarios.

3. Literature Review

3.1   Wind Power Prediction

Traditional methods (physical models, statistical)

3.2   Energy Arbitrage

Market-based strategies, storage integration

3.3   Machine Learning in Wind Energy

Common algorithms (Regression, Neural Networks, Ensemble)

3.4   Gap Analysis

What existing studies lack and how your work addresses it

4. Data and Preprocessing

Data sources (wind speed, power output, market prices)

Data cleaning and normalization

Feature selection and engineering

5. ML Models

Machine Learning (ML) is increasing using in renewable energy systems to improve forecasting and operational decision-making. In this report, ML models help predict power output and support strategies for energy arbitrage, such as deciding whether to store energy or sell it. This study focuses on two models: Regression Model, which predicts wind power output based on input parameters and Classification Regression Model, which classifies operational decisions into STORE or SELL based on input parameters. These models will be explained in detail in the following subsections.

5.1 Regression model

Before this chapter, we have already described the source of the CSV file and the location of wind farm. In this study, we implemented two regression models, which are Linear Regression and Polynomial Regression. These models were chosen because they are simple and easy to implement. It is suitable for starting models for this research.
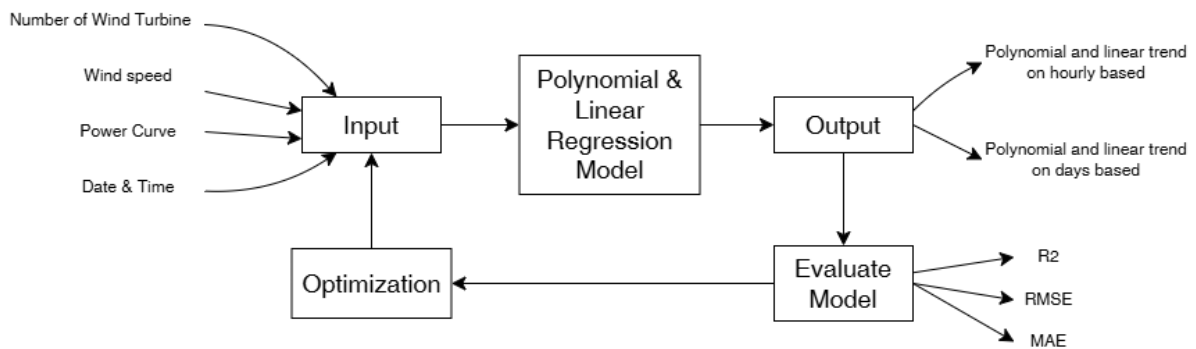

Figure 5. 1: Block Diagram of Wind farm Model (own)

In the flowchart, the modelling process considers four key input parameters are the number of wind turbines, wind speed, the power curve, and date/time information. These features were selected because they have a direct influence on wind power generation and are critical for accurate forecasting. After applying the ML models, we focused on generating two types of outputs. First, the polynomial and linear trends on an hourly basis. Second, , the polynomial and linear trends on a daily basis.

Once the models were trained and predictions obtained, we conducted an evaluation to determine which model performed best for the given dataset, which are discussed in detail in the Results section. Although optimization techniques could further improve model accuracy such as tuning or more input parameters or use different machine learning models, these were not implemented due to time constraints.

5.2   Classification Regression Model

The next model focuses on classification, specifically deciding whether to sell or store energy. For this task, we used a classification approach, and among the various available models, we selected Logistic Regression. This choice was made because Logistic Regression is simple, easy to implement, and requires fewer input parameters. The future, if the number of input parameters increases, more advanced models such as neural networks could be considered, as they offer greater predictive power for complex datasets.(give reference here)
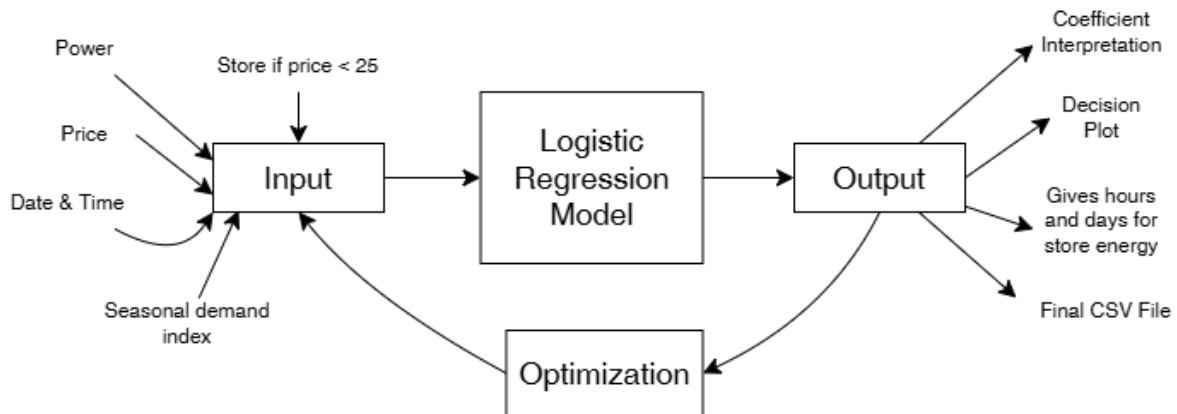

Figure 5. 2: Decision Arbitrage Model (own)

The inputs for this classification model include the predicted power on an hourly basis (obtained from the regression models), electricity price from the CSV file, date and time, and seasonal demand (assumed for this study (see table 5.1)). It is important to note that real hourly seasonal demand data in future work would improve accuracy significantly. In figure 5.2, the Logistic Regression model processes these inputs to produce several outputs. First, coefficient interpretation, which indicates the influence of each factor on the decision. Second, a decision plot showing whether to sell or store energy. Third, recommended hours and days for storing energy. Last but not least, a final CSV file containing date, hour, predicted power, price, and the store/sell decision.

Season	Months	Factor
Winter	12, 1, 2	1.2
Spring	3, 4, 5	0.9
Summer	5, 7, 8	0.8
Autumn	9, 10, 11	1

Table 5. 1: Seasonal Demand

The evaluation of this model and its performance metrics will be discussed in the Results section, while additional details and supporting figures are provided in the Appendix. Optimization techniques such as Adam optimizer or gradient descent could be applied to

improve model performance, however, due to time limitations, these were not implemented in the current work.

6. Hydrogen Storage concept

- Role of hydrogen in energy storage

- Integration with wind energy and arbitrage

- Benefits and limitations

7. Results

This section discusses the results of the wind farm model, the decision arbitrage model, and a comparison between selling without storage and using a hybrid model.

7.1   Wind Farm Model (Regression Model)

Figures 7.1 and 7.2 illustrate the prediction of wind power output. In Figure 7.1, the x-axis represents the day, and the y-axis represents power output in kilowatts (kW), which is calculated using the power curve corresponding to wind speed. The blue line shows the actual power generation on a daily basis, while the dotted green line represents the prediction trend using the Linear Regression model, and the red line shows the prediction trend using the Polynomial Regression model.
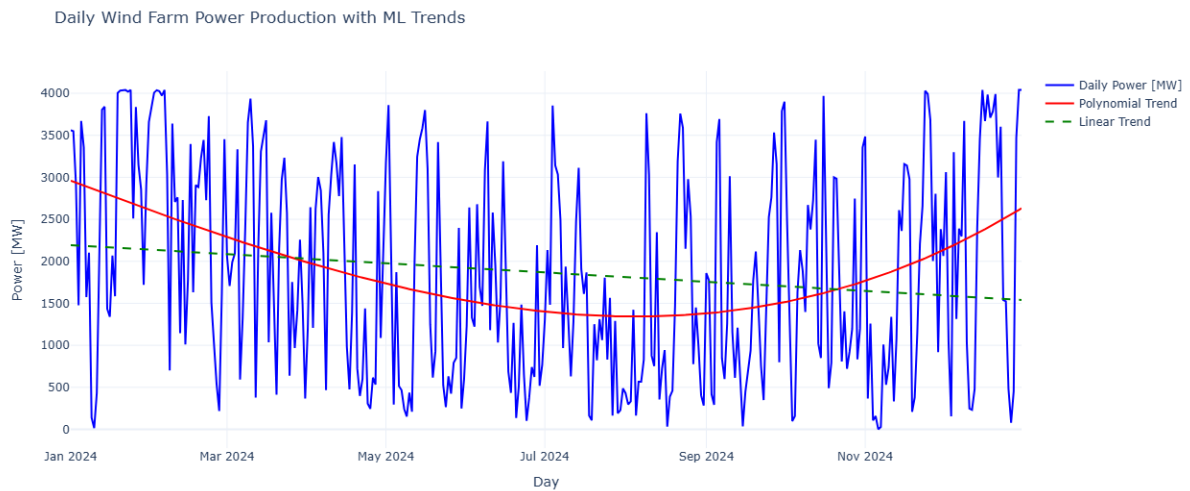

Figure 7. 1: Daily Wind Farm Power Production with ML

The evaluation metrics for these models are as follows:

- **Polynomial Regression**: R^2 = 0.14, RMSE = 1155, MAE = 993
- **Linear Regression**: R^2 = 0.02, RMSE = 1230, MAE = 1079

R^2 = Coefficient of Determination, RMSE = Root Mean Squared Error, MAE = Mean Absolute Error

Comparing these two models, the Polynomial Regression model performs better than the Linear Regression model. However, the R^2 value is still very close to zero, indicating that neither model provides strong predictive capability for daily-based data. This suggests that the models are not suitable for accurate prediction at the daily level. Therefore, in the next

section, we explore a more detailed approach by calculating power predictions on an hourly basis, which is explained in the following paragraphs.

In Figure 7.2, the x-axis represents the hours, and the y-axis represents power output in kilowatts (kW). The sky blue line shows the actual power generation on a hours basis, while the dotted green line represents the prediction trend using the Linear Regression model, and the red line shows the prediction trend using the Polynomial Regression model.
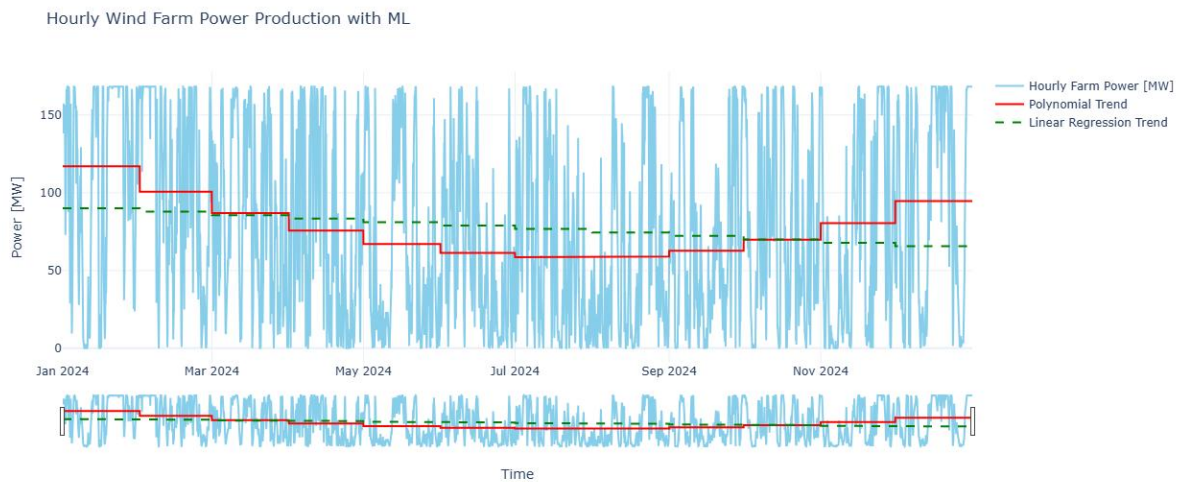

Figure 7. 2: Hourly wind Farm Power Production with ML

The evaluation metrics for these models are as follows:

- **Polynomial Regression**: R^2 = 0.94, RMSE = 3347, MAE = 2723
- **Linear Regression**: R^2 = 0.16, RMSE = 12308, MAE = 10524

Comparing these two models, the Polynomial Regression model performs significantly better than the Linear Regression model. The R^2 value for Polynomial Regression is close to 1, indicating a strong fit and good predictive capability. However, further improvements are possible by applying optimization techniques or exploring more advanced models. Additionally, introducing more input parameters in the future could enhance prediction accuracy. It is also important to note that this analysis uses only 2024 data; incorporating multiple years of data and applying advanced ML models would likely result in more accurate predictions.

7.2 Decision Arbitrage Model (Logistic Regression Model)

In Figure 7.3, the x-axis represents the hours, while the y-axis displays two aspects: the SELL or STORE decision and the electricity price. The yellow line shows the electricity spot price,

the dotted green line represents the decision to store energy, and the red line indicates the decision to sell energy to the grid.

From the figure 7.3, it is evident that most storage decisions occur during spring and summer periods. This is likely because power demand is lower during these seasons, resulting in reduced profitability when selling electricity directly to the grid. According to the model, the total number of hours designated for storage is 1147. For detailed information on the specific dates, times, and prices associated with these storage decisions, refer to the **store_hours.csv** file located in the Graphs folder of the GitHub repository.
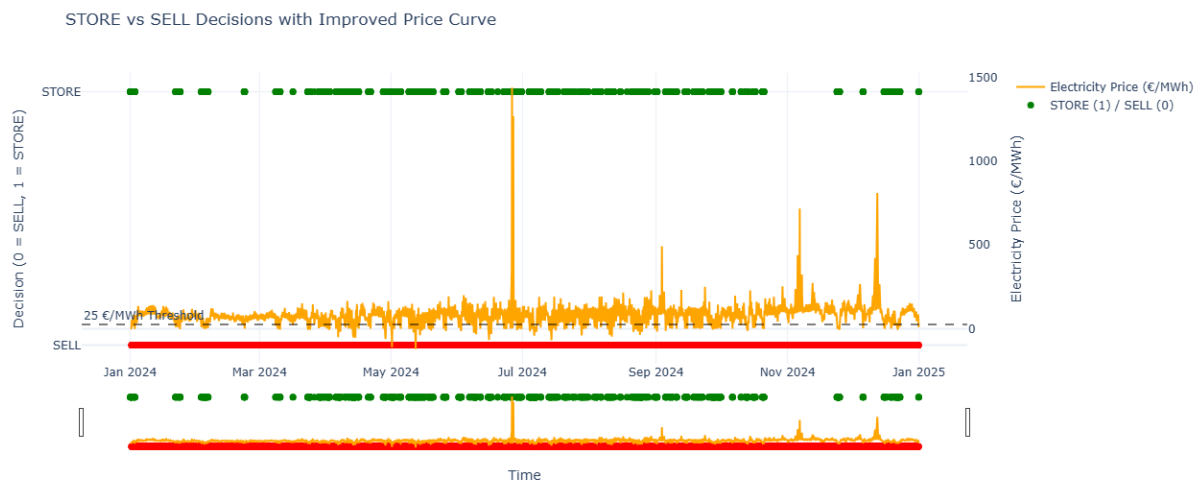

Figure 7. 3: Store Vs Sell Decisions with Improved Price Curve

Figure 7.4 presents a heatmap illustrating the STORE decision across different hours of the day. This visualization provides a more detailed view of which specific hours are recommended for storing electricity rather than selling it to the grid, primarily due to lower profitability during those periods. In the heatmap, green areas indicate hours where storing electricity is advised, while red areas represent hours where selling electricity is recommended. From the graph, it is evident that most storage decisions occur between 05:00 and 14:00, which typically corresponds to times of lower demand and reduced market prices.

From the Logistic Regression model, there were three key coefficients, Power, Price, and Seasonal Demand, with respective values of -0.01, -2.82, and -0.09. These values indicate the relative influence of each factor on the STORE or SELL decision. The negative coefficient for Price (-2.82) suggests that electricity price has the strongest impact on the decision-making process compared to the other factors. This means that when prices are low, the model is more likely to recommend storing energy rather than selling it to the grid.

While price is currently the most influential factor, it is possible to prioritize other parameters depending on operational strategies. For this study, price was considered the primary factor.

In future work, adding more parameters such as real-time demand, weather forecasts, or storage costs could lead to more advanced and accurate results.
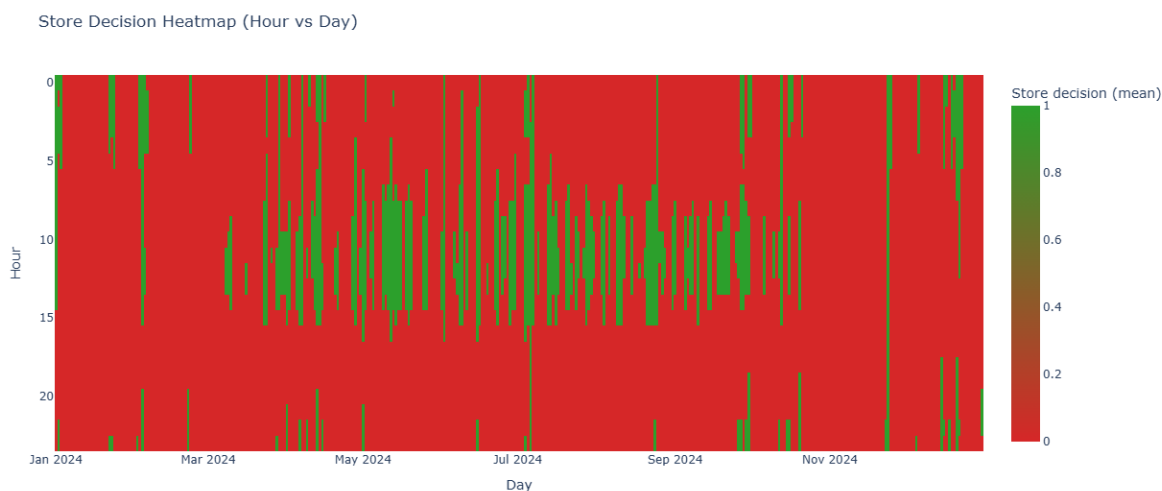

Figure 7. 4: Store Decision Hour Vs Day

7.3   Hydrogen Storage concepts

After obtaining results from Model 2, which determines whether to store or sell energy on a hours basis, we compared the net value of two scenarios, first is selling energy without storage and second is selling energy with storage. In this project, hydrogen storage was used as the primary storage system. The input parameters for the hydrogen model such as electrolyzer capacity, hydrogen production cost, and operational constraints have already been explained in the previous chapter.
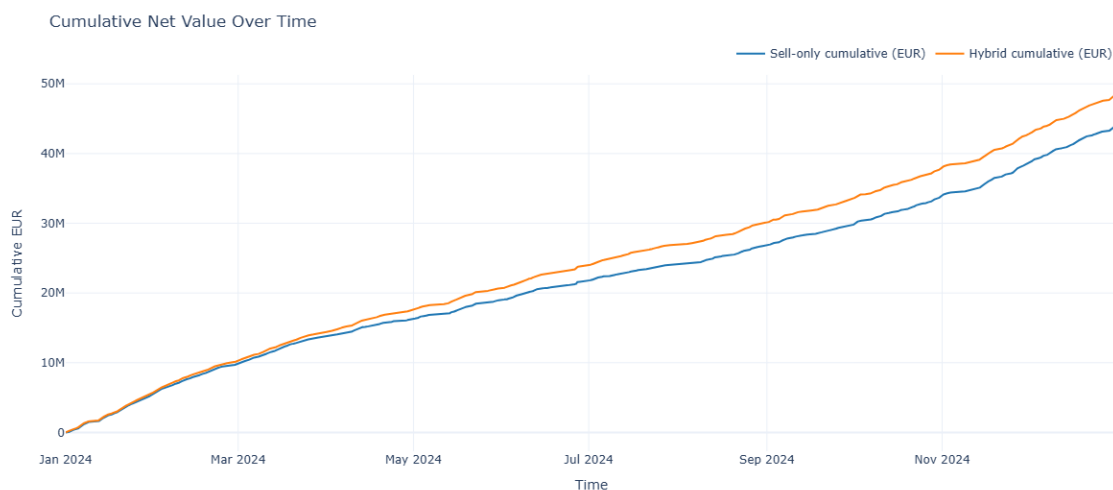

Figure 7. 5: Comparison of Net Value of With or With Storage power

Figure 7.5 illustrates the comparison of net value between two strategies, selling energy directly to the grid without storage and storing energy before selling. The x-axis represents time in hours, while the y-axis shows the net value achieved from each strategy. The blue line represents the net value when energy is sold directly to the grid without storage, and the orange line represents the net value when energy is stored and then sold based on the arbitrage strategy.

From the figure 7.5, it is clear that the storage-based strategy provides higher profitability compared to selling without storage. The final annual net value for selling without storage is approximately €44 million, whereas the net value for the storage-based strategy reaches €48.5 million. Hydrogen storage significantly improves economic performance. However, further enhancements are possible by introducing additional input parameters, using more historical data, and applying advanced optimization techniques. Due to time limitations, optimization was not the focus of this study but is recommended for future work.

8. Discussion

- Interpretation of results.

- Limitations of the study.

- Practical implications for wind farm operators.

9. Conclusion

- Summary of findings.

- Future work (e.g., real-time prediction, hybrid models).

10.   References

- All cited papers, datasets, and tools.

11.   Appendices

- Additional graphs, tables, or code snippets.

GitHub Link

On our GitHub homepage, we have attached our report, presentation, land leasing contract, and economic calculation Excel file. If you want to gain more knowledge about the topic, please check out our GitHub.