in collaboration with

# A REPORT ON

on

## Energy Forecasting


### Team: 3e-4

Manoj Khatri*

Kushal Subedi

Sunil Adhikari

{076bct034.manoj, 076bct031.kushal@, 076bei043.sunil}@pcampus.edu.np

*Author of correspondence


**Dec 25, 2023**

# Introduction

The project focuses on addressing the critical challenge of forecasting future demands and prices in the dynamic energy industry to mitigate potential financial and physical damage to resources. To tackle this issue, a resilient forecasting model is being developed, incorporating both regression and machine learning techniques. The primary objective is to empower stakeholders with accurate insights, enabling them to make strategic decisions proactively.

By leveraging regression analysis and machine learning algorithms, the model aims to provide precise predictions of future energy demands and prices. This predictive capability is crucial for minimizing risks and optimizing resource allocation in the face of the industry's inherent volatility. Furthermore, the project aligns with broader sustainability goals by fostering the transition to a net-zero future and promoting sustainable energy practices.

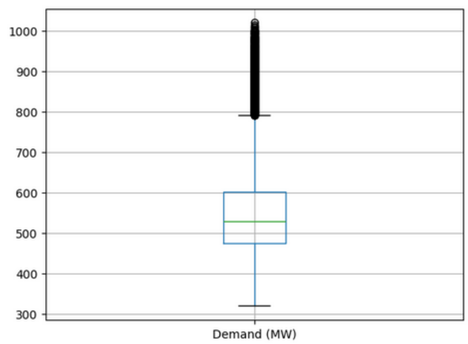
# Exploratory Data Analysis

Exploratory Data Analysis (EDA) is a crucial phase in the data analysis process, serving multifaceted purposes that extend beyond mere data exploration. It goes beyond basic summary statistics, delving into uncovering intricate patterns, trends, and outliers within the dataset. By employing statistical and visual techniques, EDA aids in identifying influential variables that significantly impact modeling outcomes. This not only refines the modeling process but also contributes to the development of more accurate and robust models. Moreover, EDA plays a pivotal role in fostering clear communication between technical and non-technical stakeholders. Through compelling visualizations and insightful summaries, it allows for the effective conveyance of complex insights, ensuring that both technical experts and non-technical decision-makers can comprehend and act upon the findings. In essence, EDA serves as the cornerstone for informed decision-making by providing a comprehensive understanding of the data landscape.

# Demand Forecasting

## Descriptive statistic of Demand:

Despite the presence of numerous outliers, they are intentionally retained in the dataset, refraining from removal. This decision stems from the recognition that these outliers might hold valuable information and convey significant insights. Removing them outright could result in the loss of crucial nuances that contribute to a more comprehensive understanding of the data. Thus, retaining these outliers aligns with the objective of capturing the entirety of the dataset's information, even if it deviates from the norm.




| | Demand (MW) |
|---|---|
| count | 27534.000000 |
| mean | 551.827235 |
| min | 320.000000 |
| 25% | 474.600000 |
| 50% | 529.500000 |
| 75% | 601.800000 |
| max | 1020.200000 |
| std | 114.424836 |

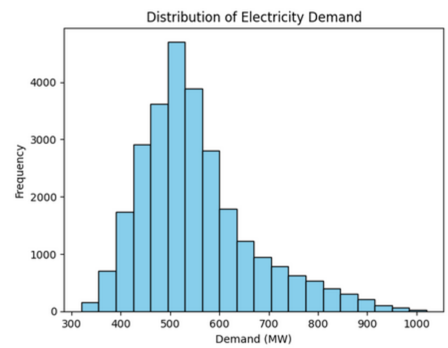



Figure 1: Descriptive statistic of Demand

**Temporal analysis of Demand:**

Demand peaks in June, July, and August, corresponding to the summer season in Pokhara, indicative of increased energy consumption, likely driven by higher usage of cooling systems during warmer months.

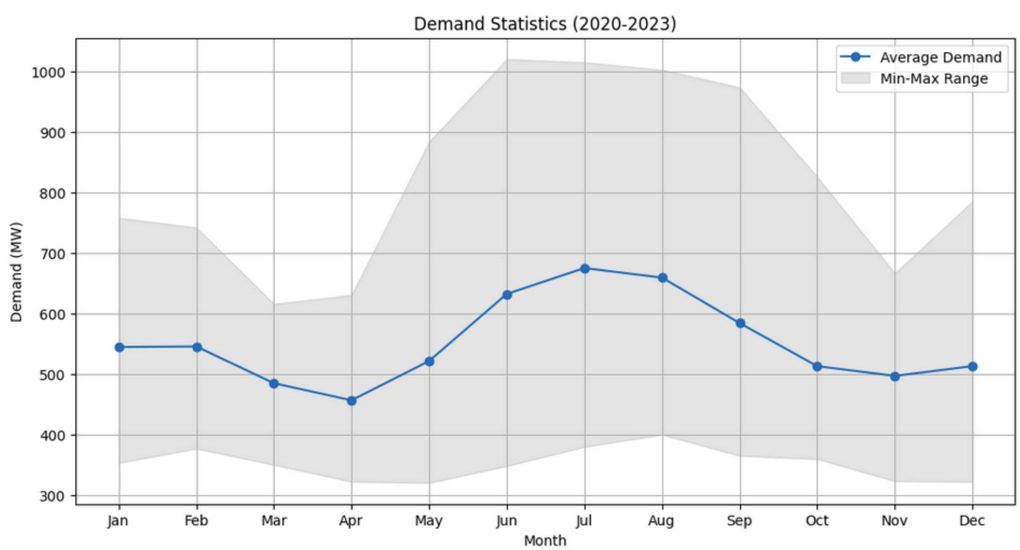



Figure 2: Temporal variation of Demand by month

The hourly demand curve exhibits an upward trend from 10 A.M. to 5 P.M., reaching its peak in the evening, notably during cooking hours.

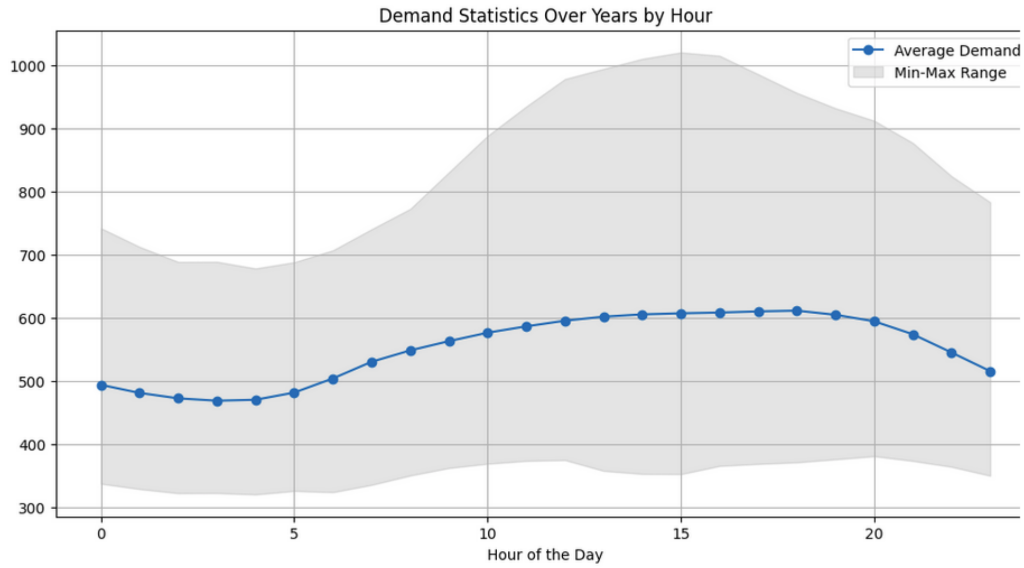



Figure 3: Temporal variation of Demand by hour

**Weather features distribution:**

The dataset comprises both normal and anomalous distributions, indicating the need for normalization techniques.

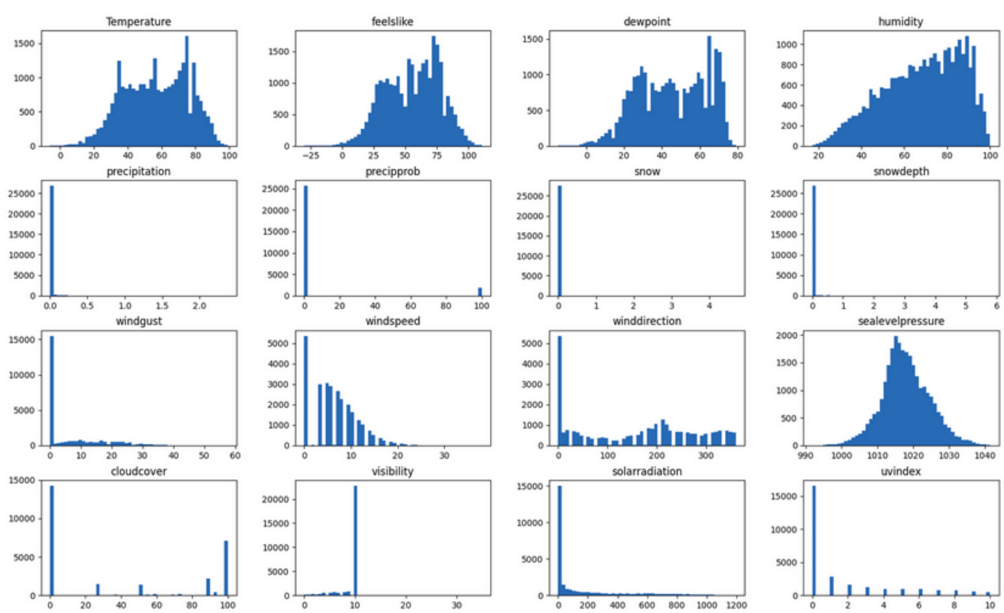



Figure 4: Histogram of weather features

**Correlation analysis of weather features:**

The provided heat map visualizes correlation coefficients among different parameters in weather data.

- Temperature, Feels-like, and Dew Point: These parameters exhibit a positive correlation, suggesting that as one of them increases, the others also tend to increase.
- UV Index and Solar Radiation: A positive correlation is observed between UV index and solar radiation. This relationship is intuitive, as higher solar radiation levels often coincide with increased UV index values.
- Wind Gust and Wind Speed: There is a positive correlation between wind gust and wind speed.

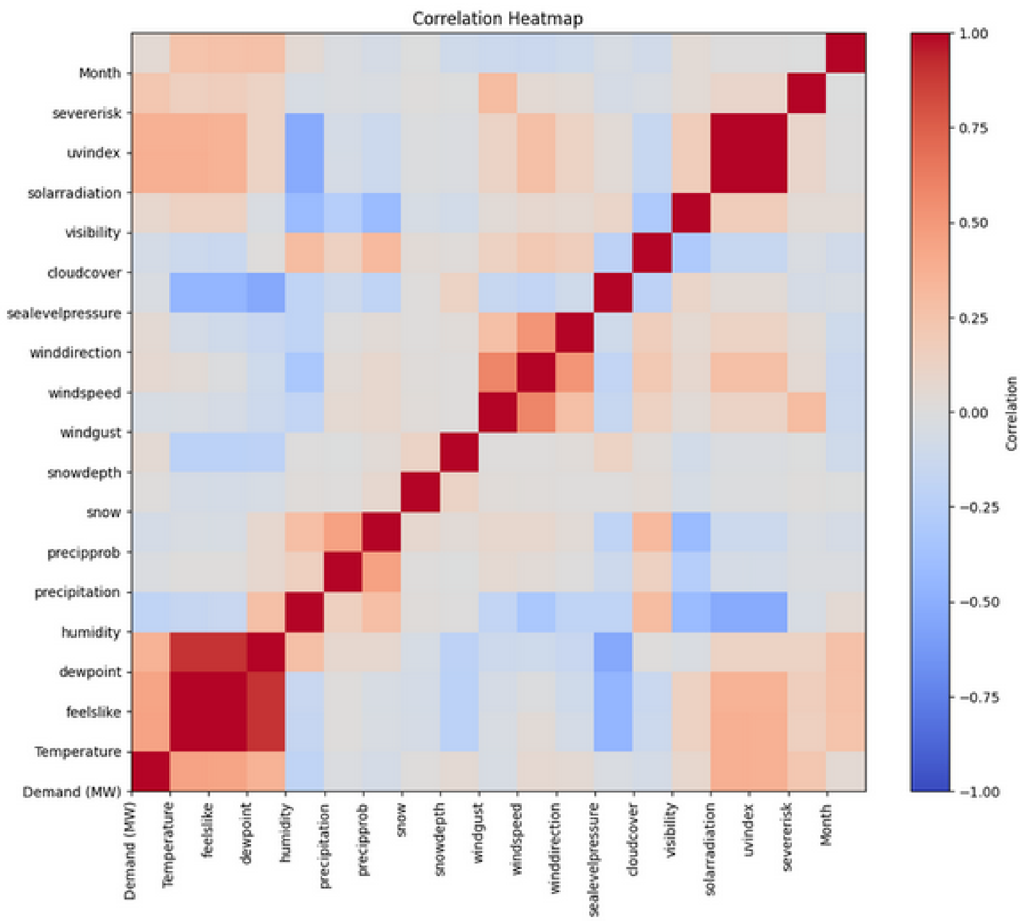



Figure 5: Heatmap of correlation between weather features

Following features seems to be highly correlated with demand:

| | |
|---|---|
| Temperature | (0.45) |
| feelslike | (0.44) |
| dewpoint | (0.36) |
| solarradiation | (0.38) |
| uvindex | (0.37) |
| severerisk | (0.52) |
| humidity | (-0.195) |

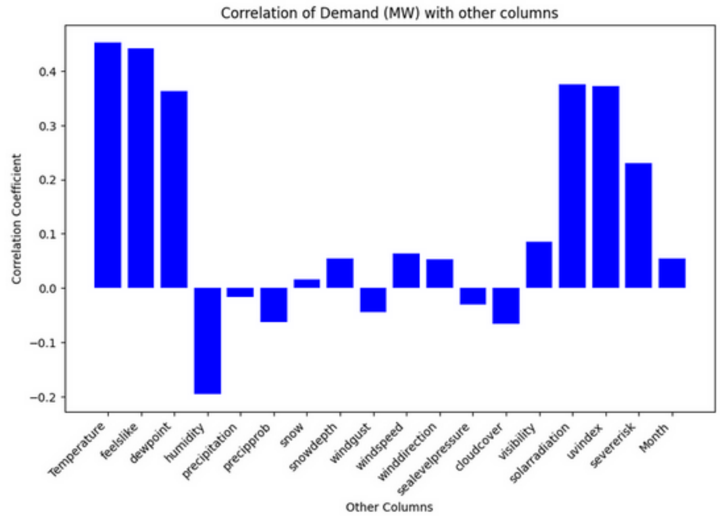



Figure 6: In depth analysis of correlation between demand and weather features

## Evaluation of Demand with respect to other weather features

The observed peak in demand during the months of June and July, corresponding to rising temperatures, indicates a positive correlation between temperature and energy demand. This relationship is often expected, as higher temperatures during the summer months typically lead to increased usage of cooling systems, such as air conditioning, contributing to heightened energy consumption.

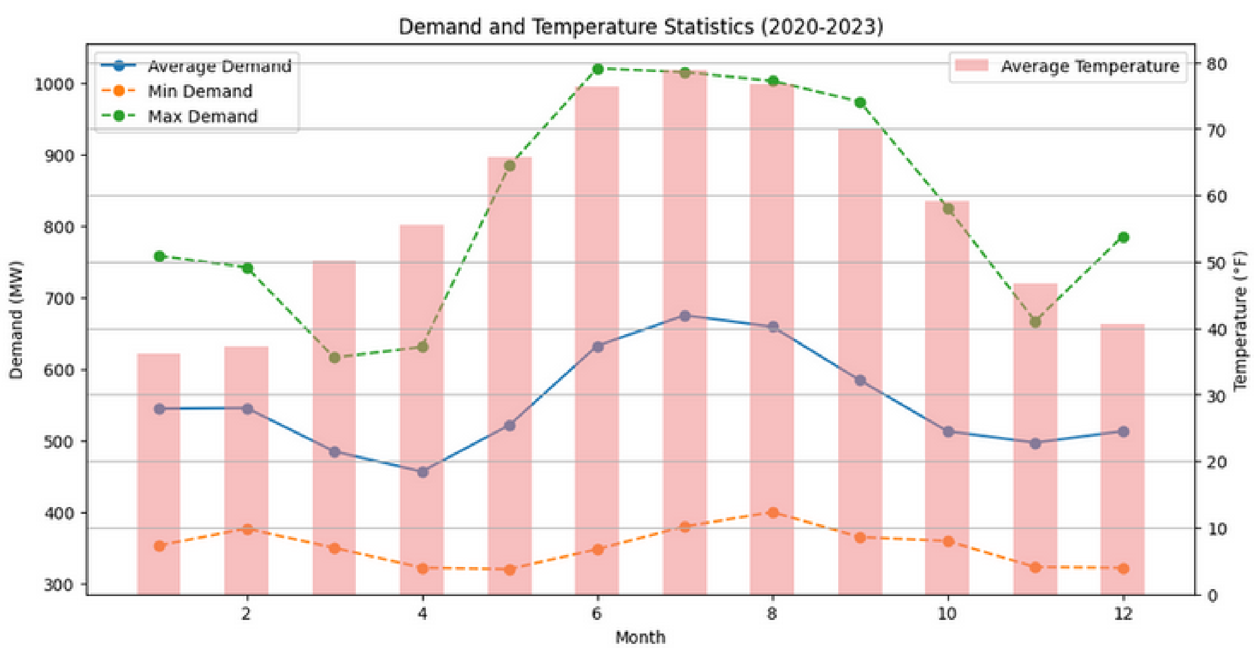



Figure 7: Demand vs temperature over months

The identified peak in demand during June and July, coinciding with increasing temperatures, suggests a positive correlation between temperature and energy demand. This relationship is in line with the common understanding that warmer weather tends to drive higher energy consumption, often attributed to the heightened use of cooling systems during hotter months.

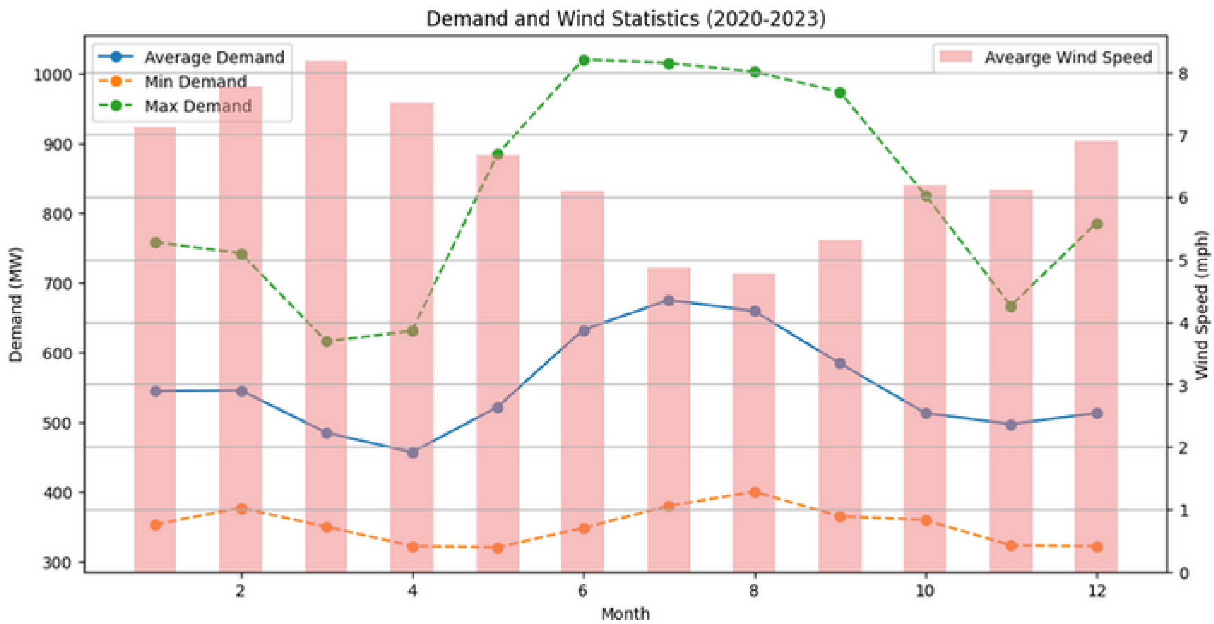



Figure 8: Demand vs windspeed over months

The observation that average electricity consumption is higher during freezing rain and snow aligns with expectations based on weather-related energy usage patterns. In colder conditions, there is an increased demand for electricity due to several factors. The operation of heating appliances becomes more prevalent as people seek to maintain comfortable indoor temperatures. Additionally, shorter days during colder seasons result in extended periods of artificial lighting use, further contributing to the overall increase in electricity consumption.

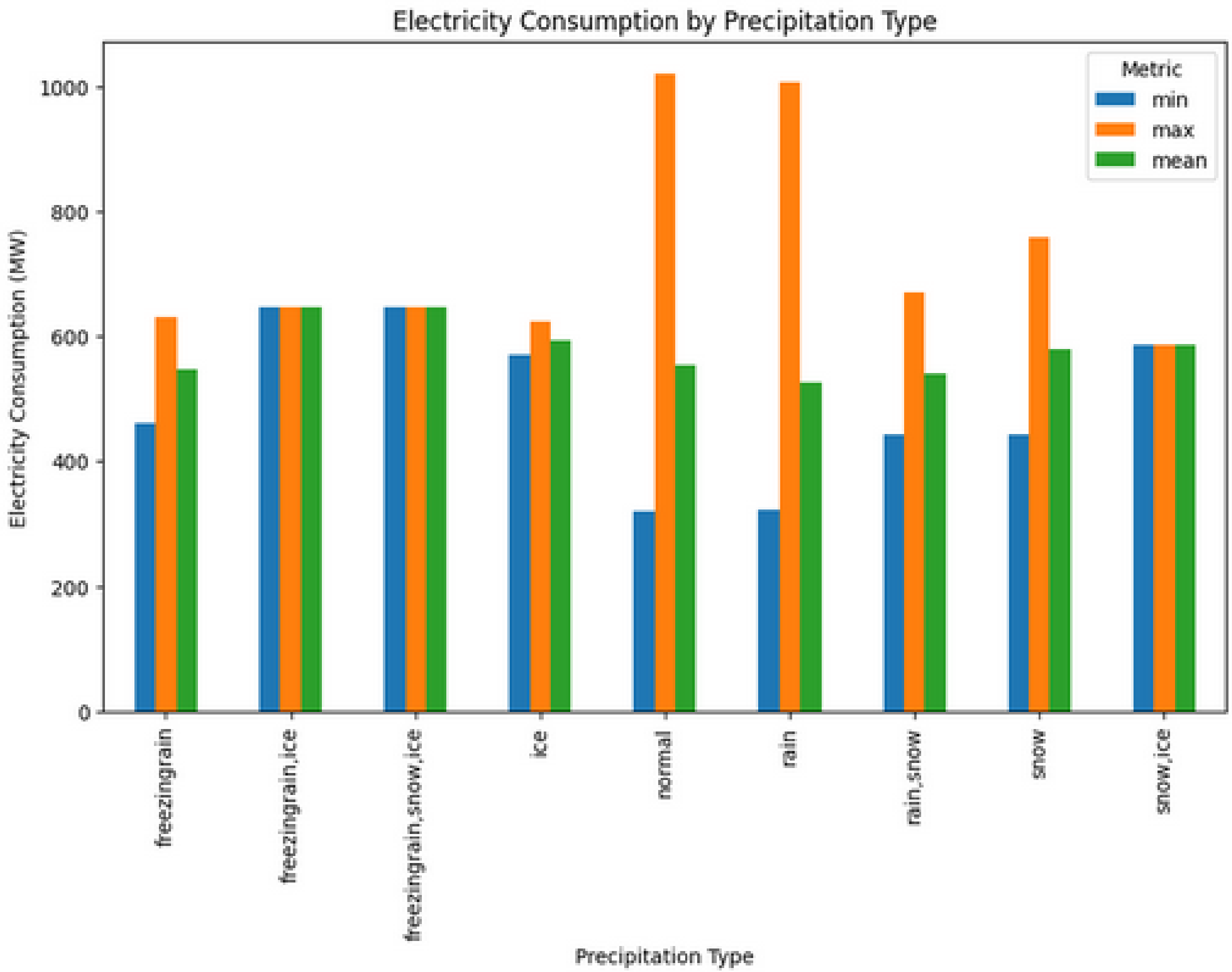



Figure 9: Demand vs preciptype

## Autocorrelation

The observed autocorrelation patterns at 1 and 24 hours imply a daily demand cycle in the dataset. This indicates that there is a recurring pattern in energy demand, with high correlation observed between values at a one-hour interval and a 24-hour interval. The consistency in these autocorrelation patterns suggests that consumer habits follow a daily rhythm, with similar energy consumption patterns occurring around the same times each day. Understanding and leveraging these patterns is crucial for accurate forecasting and optimizing resource allocation, allowing energy providers to efficiently meet the expected demand fluctuations based on consumer behavior throughout the day.

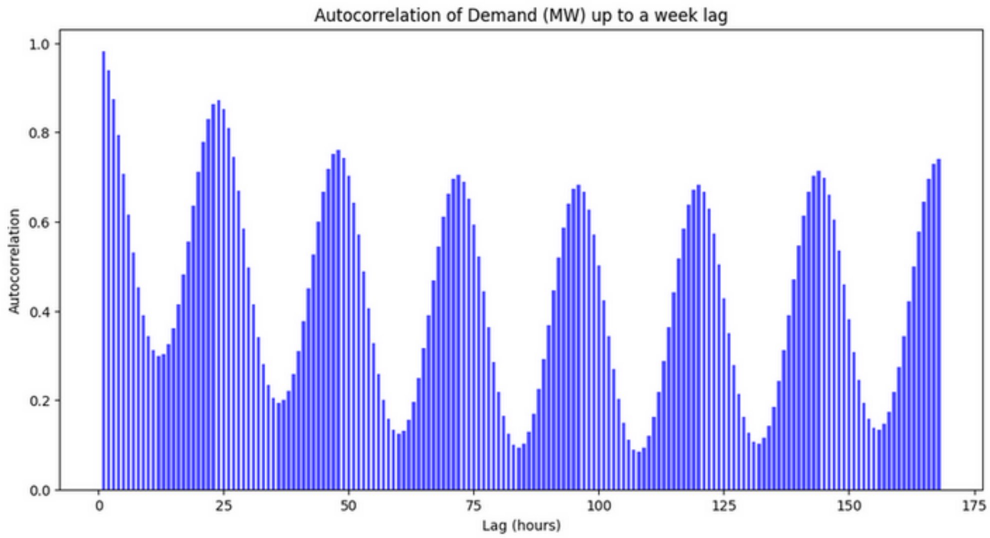
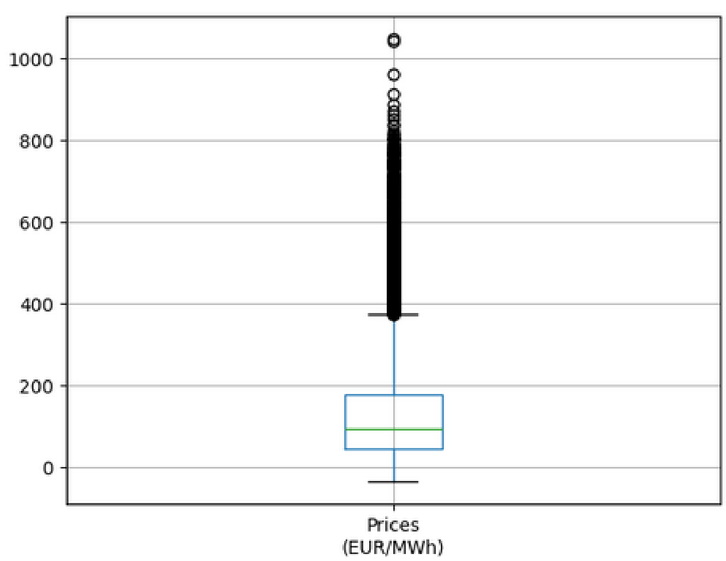



Figure 10: Demand autocorrelation by different amount of lag

# Prices Forecasting

## Descriptive statistics of Prices

The observed variation in prices, ranging from a minimum of -35 EUR to a maximum of 1047.1 EUR, reflects a diverse and potentially anomalous distribution. The presence of a negative value (-35 EUR) may indicate a peculiar occurrence or error in the data, as prices typically should not be negative.




| | Prices\n(EUR/MWh) |
|---|---|
| count | 34896.000000 |
| mean | 132.958000 |
| min | -35.000000 |
| 25% | 47.987500 |
| 50% | 95.440000 |
| 75% | 179.237500 |
| max | 1047.100000 |
| std | 120.249844 |

Figure 11: Descriptive statistic of Prices

### Temporal analysis of prices

The observed rising trend in prices from 2020 to 2022, followed by a sudden decline in 2023, suggests a notable shift in market dynamics and may indeed indicate the influence of external factors. Identifying the precise factors contributing to these fluctuations is essential for a comprehensive understanding of the pricing dynamics. Possible external influences could include changes in economic conditions, shifts in energy policies, or alterations in supply and demand patterns. Analyzing these factors in conjunction with the observed price trends can provide valuable insights for stakeholders in the industry.


Figure 12: Temporal trend of Prices

### Autocorrelation:

The autocorrelation analysis of prices reveals an interesting pattern characterized by exponential decay with oscillations. This suggests a level of persistence in price movements, wherein past prices have some influence on future ones, as indicated by the oscillations. However, the exponential decay implies that this influence diminishes over time. The observed transition into a more random pattern beyond a certain time boundary indicates that, after a certain duration, the impact of past prices on the current one becomes negligible, and prices exhibit a more stochastic behavior.

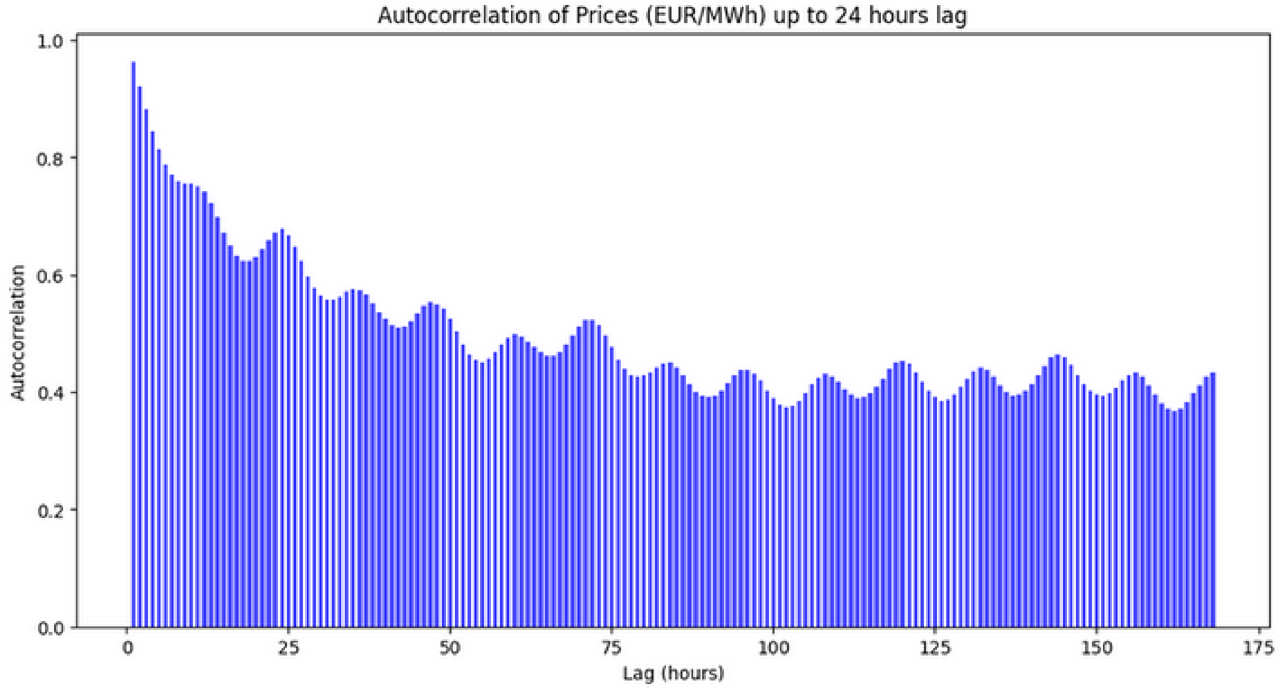



Figure 13: Autocorrelation betwen prices lag

## Feature Engineering

Feature engineering plays a pivotal role in the success of machine learning models by transforming raw data into a format that enhances the model's ability to understand underlying patterns. Effective feature engineering goes beyond algorithm selection, providing crucial insights and context to the model. Techniques such as encoding temporal information, handling categorical features, and creating new features based on domain knowledge enable models to better capture relationships within the data. By optimizing the representation of features, feature engineering contributes to model interpretability, robustness, and overall performance, ultimately influencing the accuracy and reliability of predictions in diverse applications across various domains.

The Exploratory Data Analysis (EDA) has uncovered key feature engineering techniques essential for enhancing the predictive capabilities of the model:

- **Categorical Feature Engineering**: Handling categorical features appropriately to ensure they contribute effectively to the model.

- **Wind Speed Squared and Binning Wind Direction**: Utilizing wind speed squared to capture non-linear relationships and categorizing wind direction to convey directional information.
- **Extracting Day of the Week and Month from Timestamp**: Incorporating temporal information by extracting day of the week and month from the timestamp.
- **Log Normalization**: Applying log normalization to certain features, a technique often used to manage skewed distributions and handle large value ranges.
- **Imputing NaN's with Expected Values**: Addressing missing values by imputing them with expected or calculated values through interpolation.
- **Additional Feature Columns:** Introducing new feature columns to convey supplementary information about demand, such as wind velocity squared and square root of absolute temperature.
- **Encoding Day, Hour, and Month Information**: Encoding temporal information in numerical form and adding these features to the dataframe.
- **Standard Normalization**: Applying standard normalization to standardize feature values, often crucial for certain machine learning algorithms.

These feature engineering techniques collectively contribute to a more robust and informative dataset, improving the model's ability to capture complex patterns and relationships within the data.

Additionally, interpolation techniques were applied to handle missing values in specific columns:

Linear Interpolation:

- Columns: Snow, Snowdepth, Sealevelpressure, Visibility
- Explanation: Linear interpolation was utilized to estimate missing values in these columns. This method assumes a linear relationship between consecutive data points and fills the gaps accordingly.

Spline Interpolation:

- Columns: Windgust, Severerisk
- Explanation: Spline interpolation, a more flexible method than linear

interpolation, was employed for these columns. It fits a piecewise polynomial curve to the data, providing a smoother estimation of missing values.

These interpolation techniques are valuable tools in data preprocessing, helping maintain data continuity and integrity by estimating missing values based on the observed patterns in the existing data. The choice of interpolation method depends on the nature of the data and the desired trade-off between accuracy and smoothness in the imputed values.

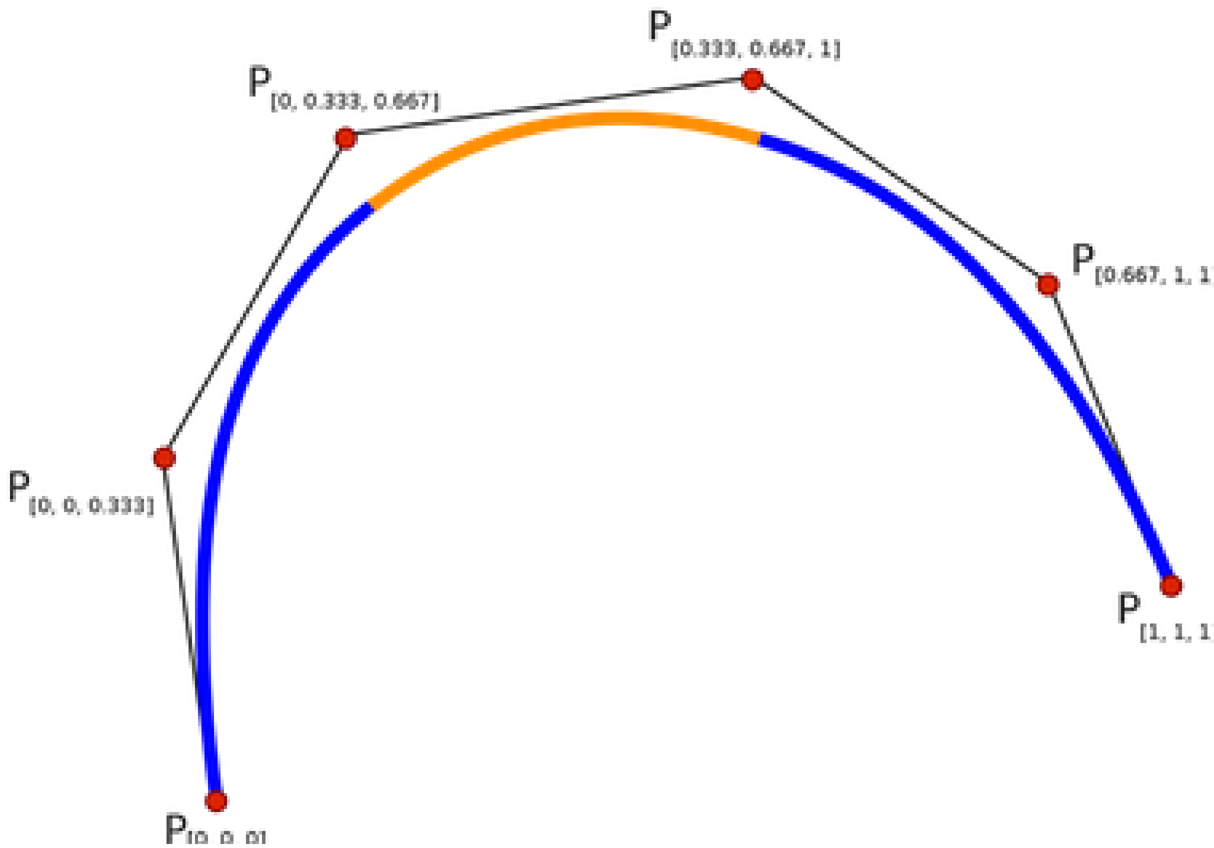



Figure 14: Spling interpolation

# Demand forecasting models

Time series data forecasting is a critical aspect of predictive analytics, finding applications in diverse fields such as finance, weather forecasting, and demand planning. The primary goal of forecasting models is to predict future values based on historical observations, allowing organizations to make informed decisions and allocate resources efficiently. These models leverage various techniques, including classical statistical methods, machine learning algorithms, and neural networks.

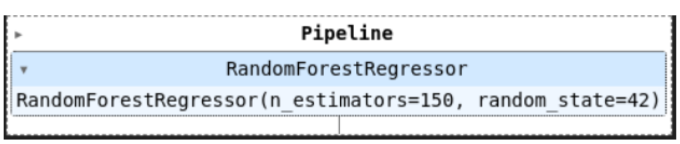
## Classical ML

Classical machine learning (ML) techniques, while effective for near-term predictions, often encounter challenges when applied to longer time frames. One significant limitation is the difficulty in tracking model convergence over extended periods. Classical ML algorithms may struggle to adapt to evolving patterns or trends that emerge over time, leading to suboptimal performance for forecasting tasks with longer horizons.

**Disadvantages:**

1. Pruning, a common technique to enhance model efficiency, can compromise the model's capability when applied over extended time frames. As data distributions evolve, the features deemed irrelevant during pruning might become crucial for accurate predictions in the future.

2. Another concern arises with large datasets, as classical ML techniques may face difficulties in processing and extracting meaningful insights from vast amounts of data.



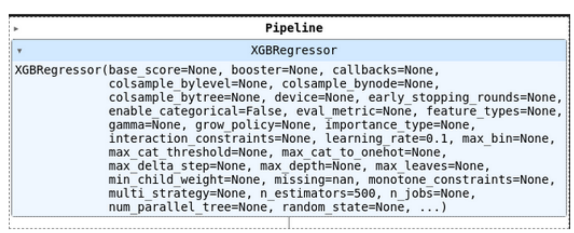




Figure 15: Our implementation of classical ML techniques

### Deep Learning

Neural networks offer several advantages over classical machine learning techniques, making them particularly powerful in handling complex tasks and large datasets:

- Neural networks handle non-linear relationships more effectively, capturing complex patterns in data.
- They automatically learn and extract features from raw data, reducing the need for manual feature engineering.
- Specialized architectures like RNNs and LSTMs excel at capturing temporal dependencies in sequential data.
- Neural networks scale efficiently to handle large datasets and adapt to changing patterns, making them suitable for dynamic environments.

Our particular reasons for using Neural Networks are:

1. Number of data points » Dimension of data.
2. Deep networks suitable for time series inductive prior.
3. Scalable for large datasets.

The neural network architecture comprises five hidden layers with dimensions decreasing from 256 to 1, employing the Rectified Linear Unit (ReLU) activation function. This design indicates a progressive reduction in the learned representations, potentially facilitating the extraction of hierarchical features. The model encompasses a total of 98,049 parameters, reflecting its capacity to adapt and capture intricate patterns in the data during the training process.

| **Model Architecture** | |
|---|---|
| No of Hidden Layers | 5 |
| Hidden Dimensions | $256 \to 128 \to 64 \to 32 \to 1$ |
| Activation Function | ReLU |
| Number of Parameters | 98,049 |

Table 1: Details of the Model Architecture

The training parameters for the neural network model include a learning rate of $3 \times 10^{-4}$, utilizing the Adam optimizer for gradient descent optimization, and employing Tensorboard for tracking. The model is trained for approximately

100 epochs. The resulting mean squared error (MSE) metrics reveal a training MSE of approximately 119.43 and a validation MSE of around 153.34, computed on an 80-20 split for training and validation data. It's crucial to note that MSE, while commonly used for performance evaluation, can be deceptive, and interpretation should be done with consideration of the specific context and characteristics of the data.

| **Training Parameters** | |
|---|---|
| Learning Rate | $3 \times 10^{-4}$ |
| Optimizer | Adam |
| Tracker | Tensorboard |
| Epoch | ∼100 |
| **Performance Metrics** | |
| Training MSE | ∼119.43 |
| Validation MSE | ∼153.34 |
| **Note** | |
| * 80-20 split used for training and validation.<br>** MSE can be deceptive. | |

Table 2: Training Parameters and Performance Metrics

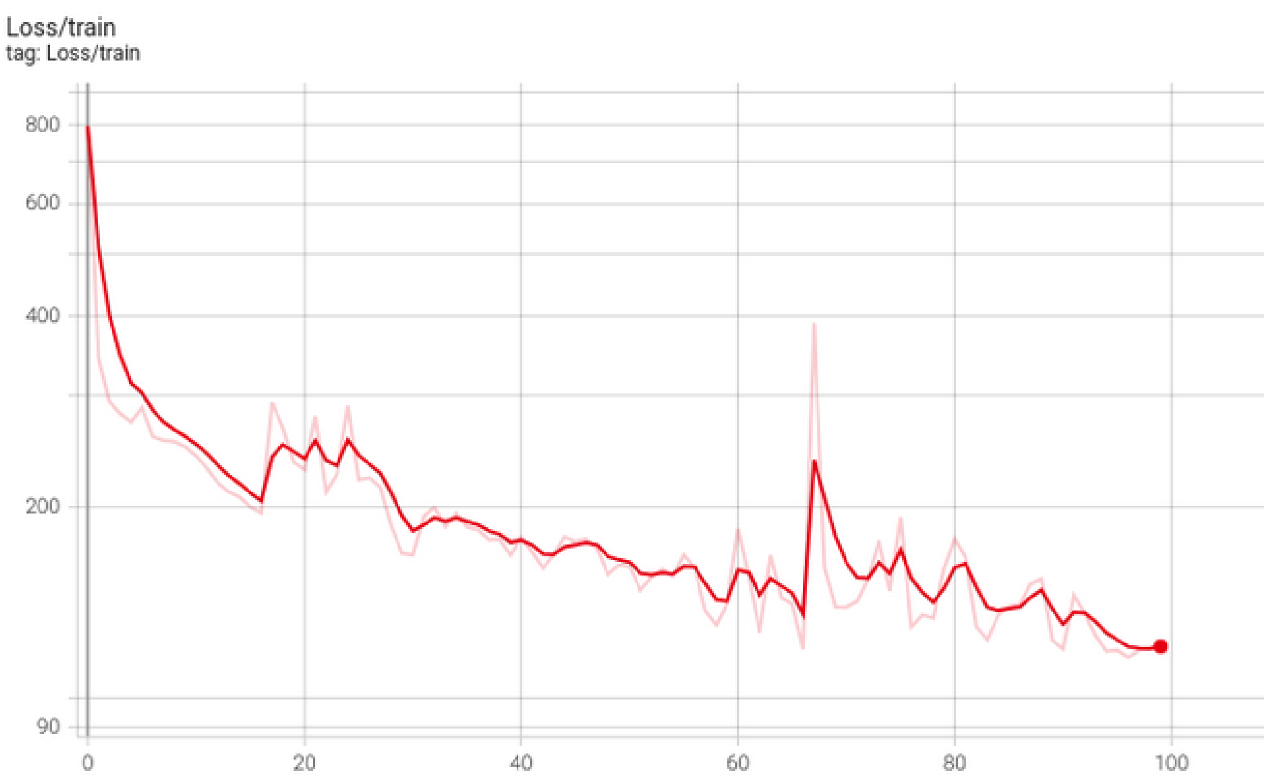



Figure 16: The training curve of the NN model

### Model Performance

Our model exhibits a notable capability to capture both short-term and long-term dependencies within the data. The architecture's depth, featuring multiple hidden layers, enables the model to discern intricate patterns over different time scales.

The decreasing dimensions in the hidden layers, coupled with the ReLU activation function, facilitate the extraction of features at various levels of abstraction, contributing to the model's adaptability in understanding both immediate and prolonged relationships within the input data. This dual proficiency in capturing short-term nuances and long-term trends enhances the model's versatility across a range of temporal patterns and contributes to its robust predictive performance.

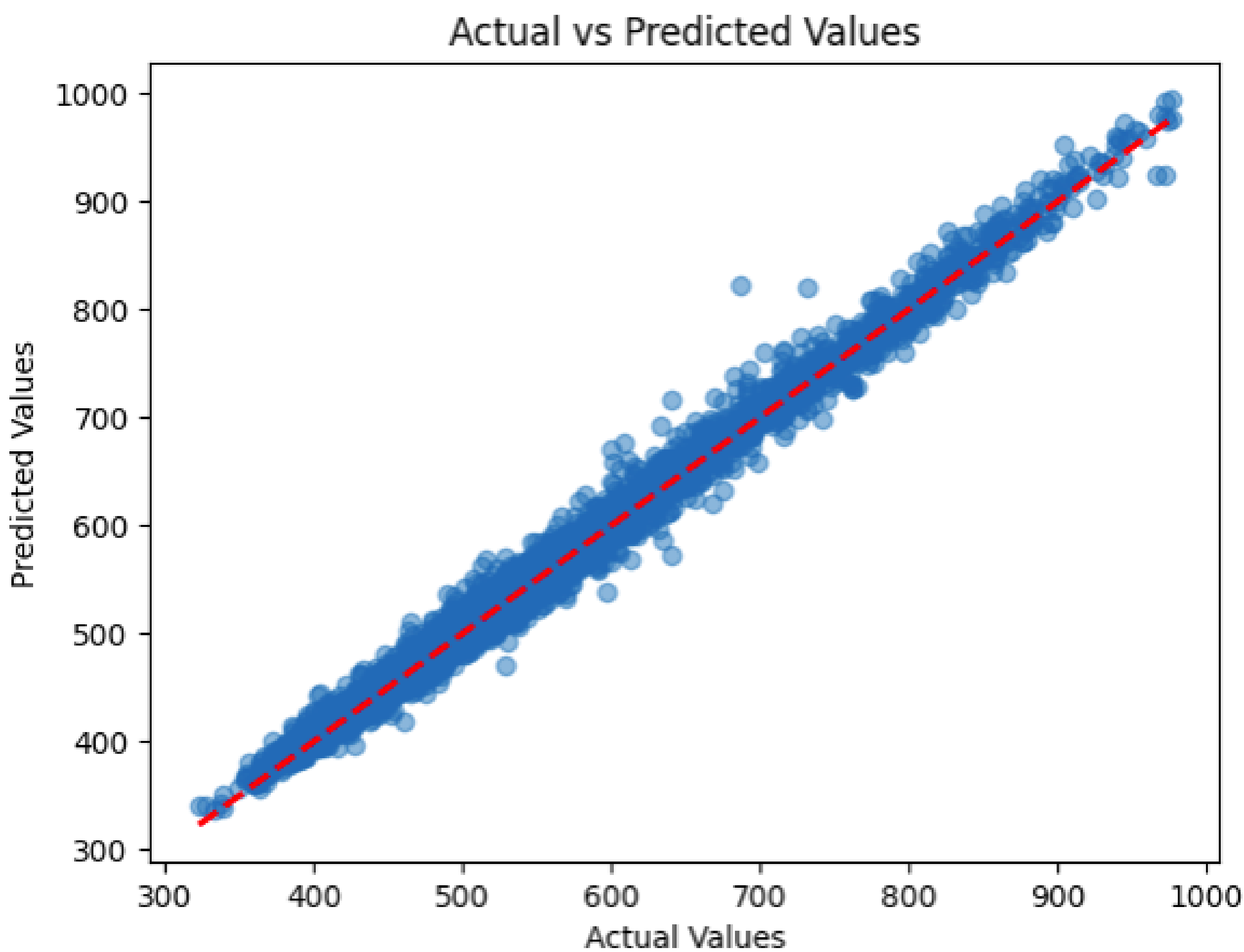



Figure 17: Short term prediction of our model in train set

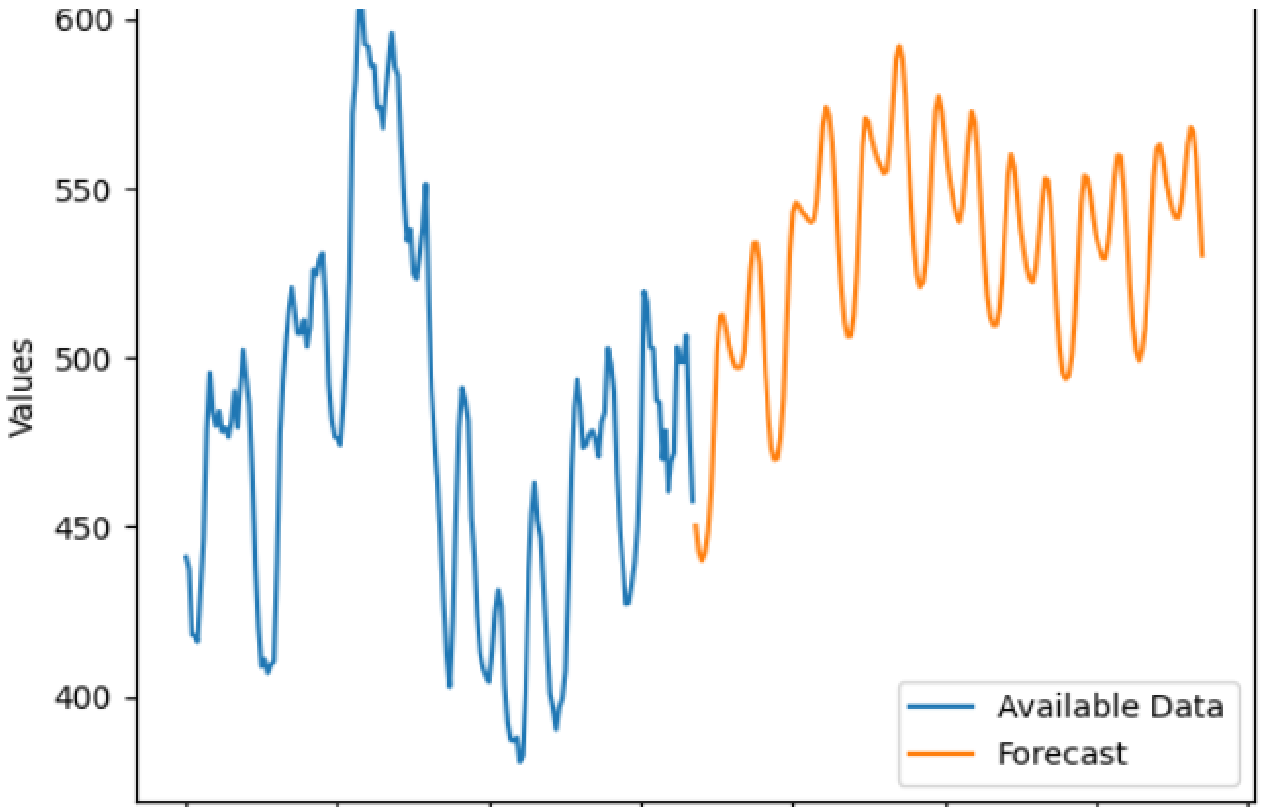



Figure 18: Long term prediction of our model for 7 days (capures hourly and daily patterns)

## Prices forecasting models

In the context of price forecasting, the utilization of Bidirectional Long Short-Term Memory (BiLSTM) models has demonstrated significant advantages over unidirectional LSTM counterparts. A analysis (Siami-Namini et. al) reveals a remarkable 37.78% reduction in error rates when employing BiLSTM for forecasting time-series data, particularly in the dynamic realm of price fluctuations.

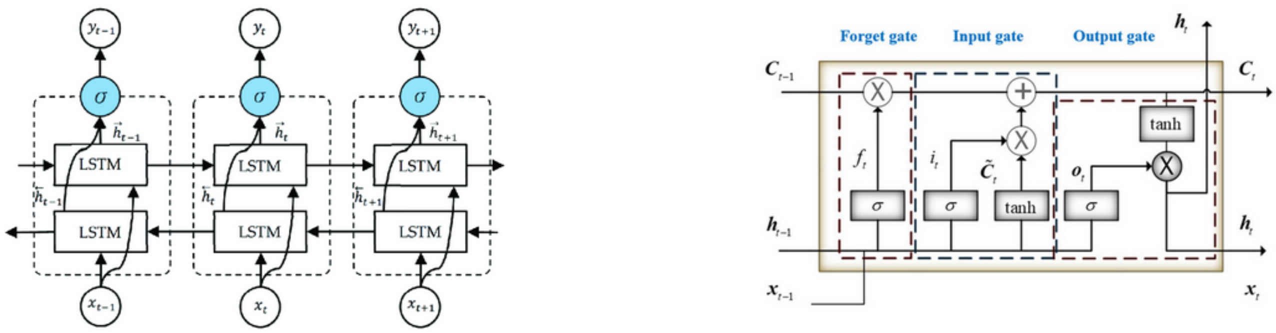



Figure 19: BiLSTM and LSTM unit)

BiLSTM's distinctive feature lies in its ability to capture essential additional features by processing input sequences in both forward and backward directions, enhancing its understanding of complex temporal patterns. This bidirectional approach allows the model to consider context from both past and future observations, providing a more comprehensive perspective on the underlying dynamics of price movements.

**Comparative analysis of LSTM vs BiLSTM on Hitachi prices dataset.**

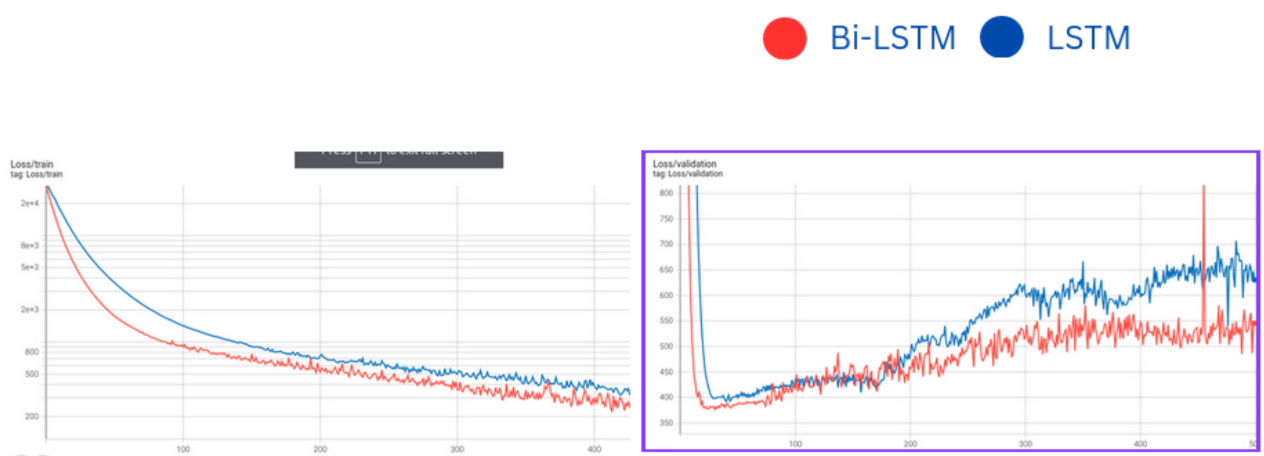



Figure 20: Comparative training of LSTM vs BiLSTM on prices dataset of Hitachi (BiLSTM is stable)

The model architecture employed for time series forecasting involves a Bidirectional Long Short-Term Memory (BiLSTM) network. This architecture is characterized by its ability to capture both past and future context, contributing to a more comprehensive understanding of temporal patterns within the data. The key architectural parameters are set as follows: an input size of 1, a hidden size of 50, and a stacking of 5 BiLSTM layers. The total number of parameters in this configuration is 264,501.

For the training process, the time series data is prepared using a lookback of 168 time steps, equivalent to 7 days, and organized into batches of 32 samples. The learning rate is set to $3 \times 10^{-4}$, and the model undergoes training for 500 epochs. During training, the model learns to predict future values by iteratively adjusting its parameters based on the provided historical data. The objective is to minimize the mean squared error (MSE) between the predicted and actual values. The BiLSTM architecture, coupled with the specified training parameters, aims to capture and leverage both short-term and long-term dependencies in the time series data, enhancing its ability to make accurate predictions. The training process involves iterative optimization, where the model refines its internal representations to improve forecasting performance over time.

| **BiLSTM Architecture** | |
|---|---|
| Input Size | 1 |
| Hidden Size | 50 |
| Number of Stacked Layers | 5 |
| Total Parameters | 2,64,501 |

Table 3: BiLSTM Architecture Parameters

| **Training Parameters** | |
|---|---|
| Lookback | 168 (7 days) |
| Batch Size | 32 |
| Learning Rate | $3 \times 10^{-4}$ |
| Number of Epochs | 500 |

Table 4: Training Parameters for Time Series Forecasting

## Model Performance:

The model's performance on the test data, as indicated by a mean squared error (MSE) of approximately 376, suggests a favorable alignment with the current trend in the data. The MSE is a measure of the average squared difference between predicted and actual values, and a lower MSE indicates better accuracy.

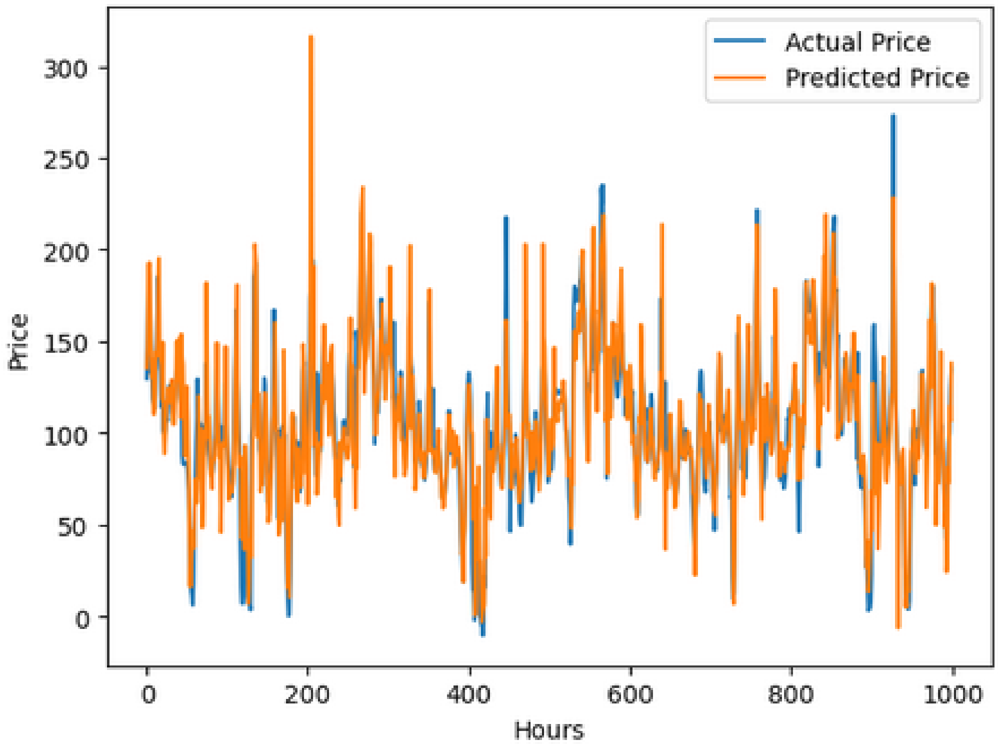



Figure 21: Model performance on test set

After training on 80% of the dataset, the application of the BiLSTM model for forecasting the next 7 days demonstrates its capability to accurately capture the overall trend, albeit with some slight deviations. The graph visually showcases the model's proficiency in understanding and replicating the underlying patterns present in the time series data.

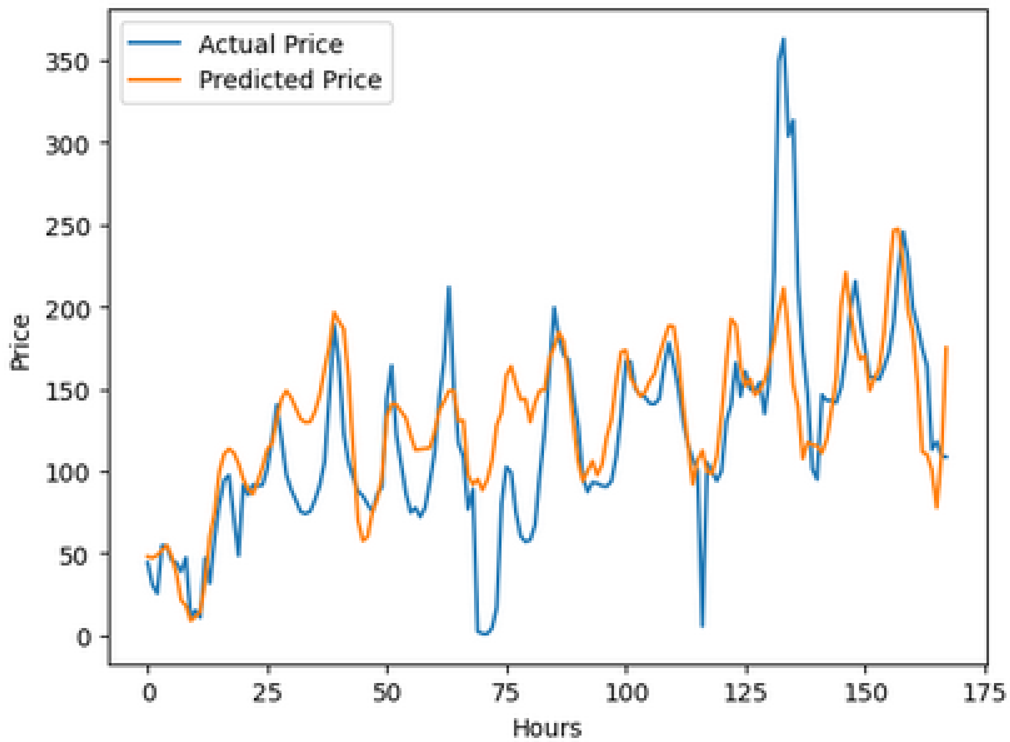



Figure 22: Model's long term predictive capabilities

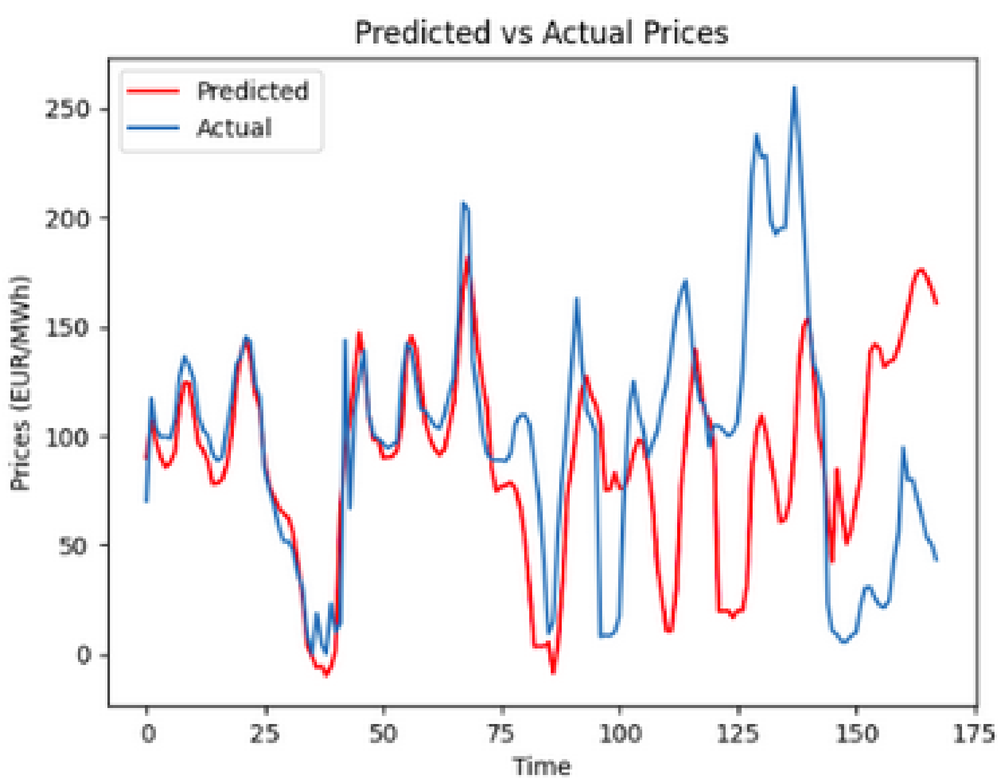



Figure 23: The model subtely captures the pattern upto 3 days then there is a bit divergence.

**Final Forecasting** The ultimate model, trained on the complete dataset and yielding a Mean Squared Error (MSE) of approximately 27, predicted the prices for the upcoming 7 days with an hourly granularity, as illustrated below.

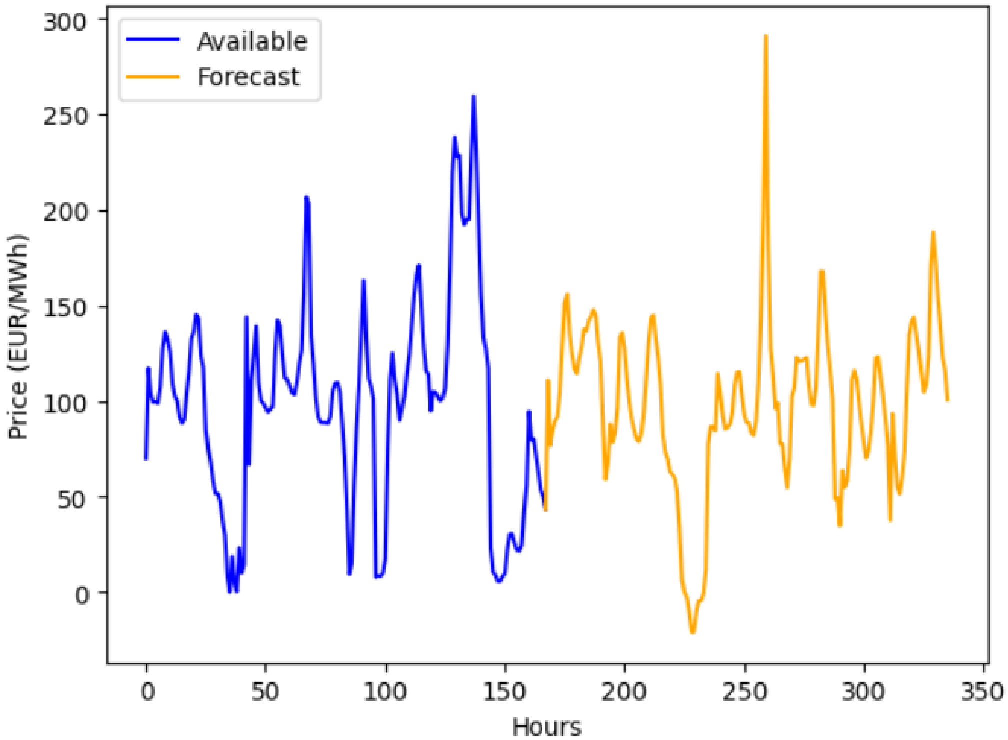



Figure 24: The model is able to capture dropping and sudden increase within the dataset

## Conclusion

In conclusion, the integration of machine learning techniques plays a pivotal role in forecasting future demands and prices, acting as a proactive measure against potential financial or physical resource damage. The adaptability of our models ensures resilience in dynamic market conditions, allowing for timely adjustments and responses to evolving trends. The informed decision-making facilitated by accurate predictions leads to significant cost savings and strategic resource management.

**We hope we could play our part towards Hitachi's mission for sustainability.**

# References


- Siami-Namini, S., Tavakoli, N., & Siami Namin, A. (2019). Analysis. In Proceedings of the IEEE International Conference on Big Data.