Solution Report

Preprocessing of Data

The preprocessing method used in preparing the data for the prediction model involved several steps to ensure the data's quality and suitability for training the LSTM neural networks. Here's an outline of the preprocessing method:

1. Data Loading: The dataset containing historical average prices (EUR/MWh) data was loaded into a panda DataFrame.

2. Handling Missing Values: Any missing values in the dataset were identified and removed using the ‘**dropna()**’ function, ensuring that only complete data points were used for training and testing.

3. Identifying Outliers:
   - Outliers were detected using statistical methods such as calculating the lower and upper quantiles.
   - Threshold values for the bottom 1% and top 1% of the data were determined using the ‘**quantile()**’ function.
   - Data points falling below the bottom 1% or above the top 1% thresholds were considered outliers and replaced with the average of the neighboring values.

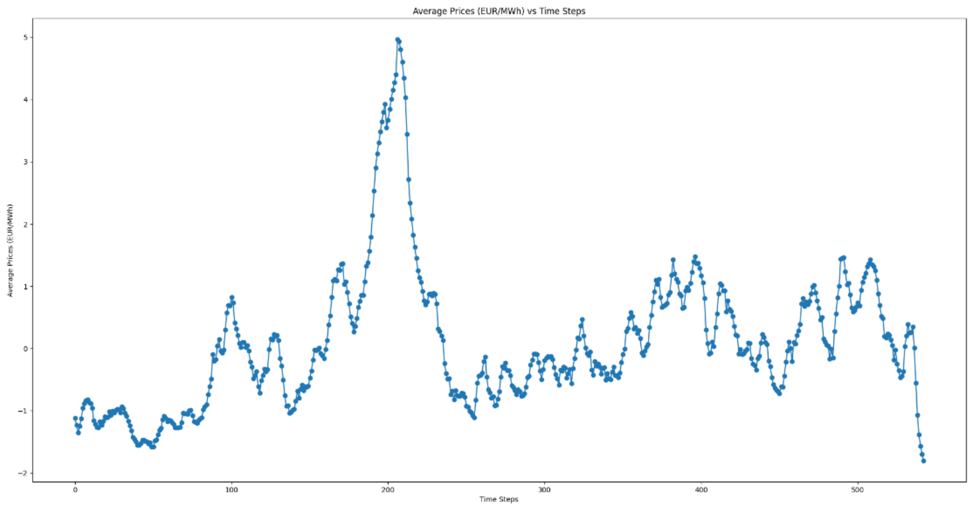
4. Smoothing Data:
   - A rolling window approach was applied to smooth the data and remove noise. This involved calculating the moving average over a specified window size.
   - The ‘**rolling()**’ function in pandas was used to compute the rolling mean over the dataset.

5. Standardization:
   - The data was standardized to have a mean of 0 and a standard deviation of 1 to ensure that all features were on a similar scale.
   - The ‘**StandardScaler**’ from the ‘**sklearn.preprocessing**’ module was used to perform standardization on the data.

6. Data Transformation:
   - The data was transformed into a format suitable for training the LSTM model. This involved reshaping the data into sequences of input-output pairs.
   - Sequences of past data points (input sequences) were used to predict the subsequent data point (output sequence).
   - The input sequences were created by shifting a window of past data points.
   - The ‘torch.FloatTensor()’ function was used to convert the data into PyTorch tensors, and the ‘unsqueeze()’ function was used to add an additional dimension to the data.

7. Splitting Data:
- The preprocessed data was split into training and testing sets, with 80-20 ratio.
- The ‘train_test_split’ function from ‘sklearn.model_selection’ module was used for this purpose.

By following these preprocessing steps, the dataset was cleaned, normalized, and transformed into a format suitable for training the LSTM neural networks for time series forecasting. This ensured that the models were trained on high-quality data, leading to more accurate predictions.



Summary of the Model used:

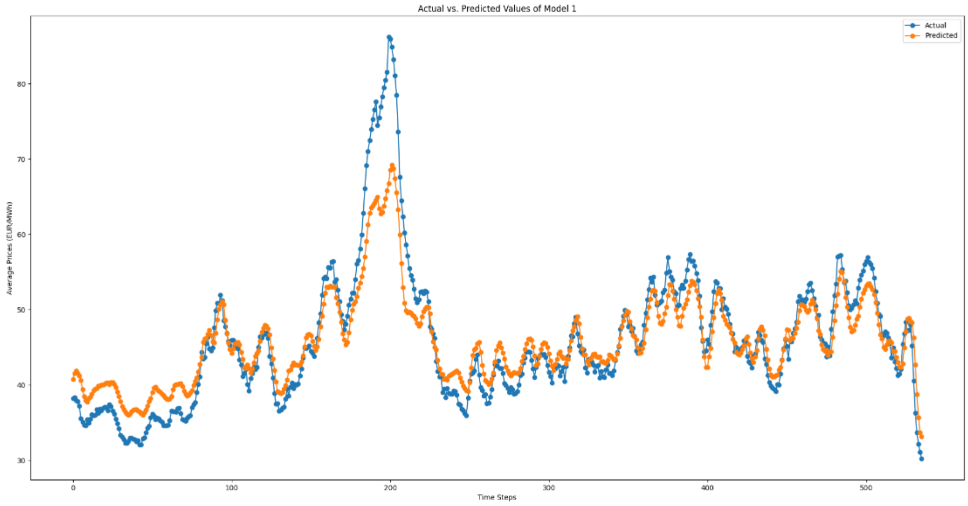
This report presents the development and evaluation of a prediction model using LSTM neural networks for time series forecasting. The goal of the project was to predict average prices (EUR/MWh) based on historical data. The dataset was pre-processed to handle missing values and outliers, and then split into training and testing sets. Three different LSTM models were implemented and trained using the training data. The models were evaluated using the testing data, and the best-performing model was selected based on its accuracy.

Model Architectures:

In all the models context window has been taken as 7 days

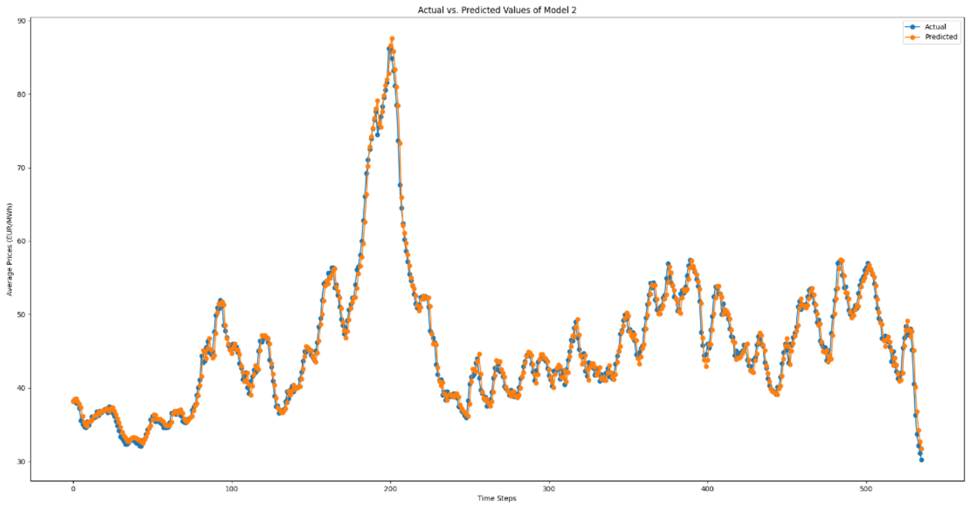
1. Improved LSTM:

- Architecture: 2 LSTM layers with batch normalization and dropout
- Hyperparameters: Input size = 1, hidden size1 = 128, hidden size2 = 128, output size = 1
- Training: Adam optimizer, MSE loss function



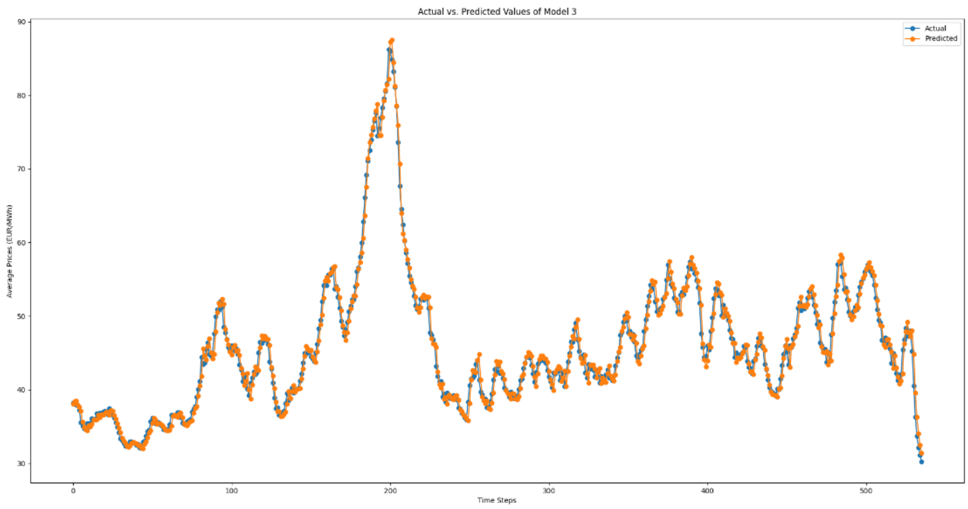
2. Better LSTM:

- Architecture: 2 LSTM cells with linear layers and dropout
- Hyperparameters: input size = 1, hidden size1 = 256, hidden size2 = 128, output size = 1, dropout rate = 0.25
- Training: Adam optimizer, MSE loss function



3. Another LSTM (Model3):

- Architecture: Similar to Better LSTM with different hidden sizes
- Hyperparameters: input size = 1, hidden size1 = 256, hidden size2 = 256, output size = 1, dropout rate = 0.25
- Training: Adam optimizer, MSE loss function



Evaluation:

The models were evaluated using the testing data, and the performance was measured using the Adjusted R-squared metric. Adjusted R-squared quantifies the proportion of the variance in the dependent variable that is predictable from the independent variables. A higher Adjusted R-squared value indicates a better fit of the model to the data.

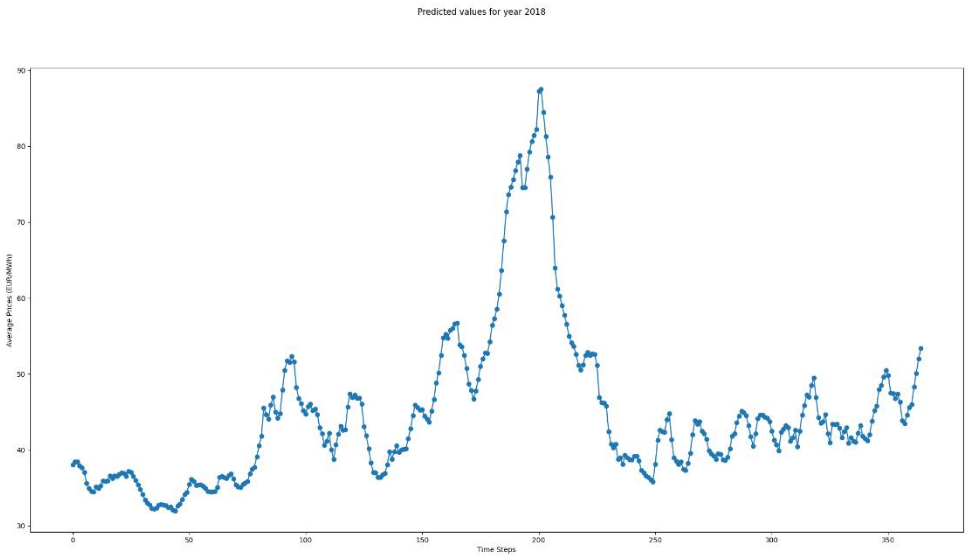
Accuracy Score (Adjusted R-squared):

- Improved LSTM: 0.9686
- Better LSTM: 0.9722
- Another LSTM (Model3): 0.9843

Conclusion:

Based on the evaluation results, Another LSTM (Model3) achieved the highest accuracy with an Adjusted R-squared score of **98.43%**. This model demonstrated superior performance in predicting average prices, suggesting its effectiveness for time series forecasting tasks.

This concludes the Solution Report summarizing the prediction model development and evaluation process.



This is the predicted daily average price (EUR/MWh) of the year 2018

<u>Optimization</u>:

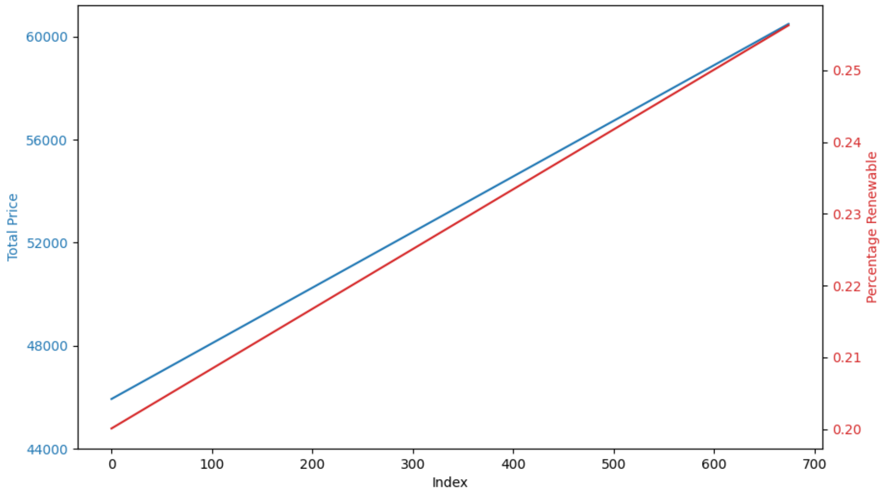
Observation:

As the quantity of electricity drawn from the State Electricity Grid (Q_Grid) increases, there is a noticeable positive correlation with the percentage of total renewable electricity. This indicates that increasing Q_Grid tends to lead to a higher proportion of renewable energy in the total electricity consumption. However, this relationship is also accompanied by an increase in the total price of electricity.

Challenges:

The code provided plots the total price and the percentage of renewable electricity against the index (representing different values of Q_Grid and Q_Exchange). Both plots exhibit almost the same positive slope, indicating that it is challenging to decrease one variable while increasing the other.

Analysis:



The graph illustrates the trade-off between the quantity of electricity drawn from the State Electricity Grid and the proportion of renewable energy in the total electricity consumption. While higher Q_Grid values result in a greater percentage of renewable energy, they also lead to higher total prices. This trade-off poses a challenge in finding an optimal solution that maximizes renewable energy usage while minimizing costs.

Conclusion:

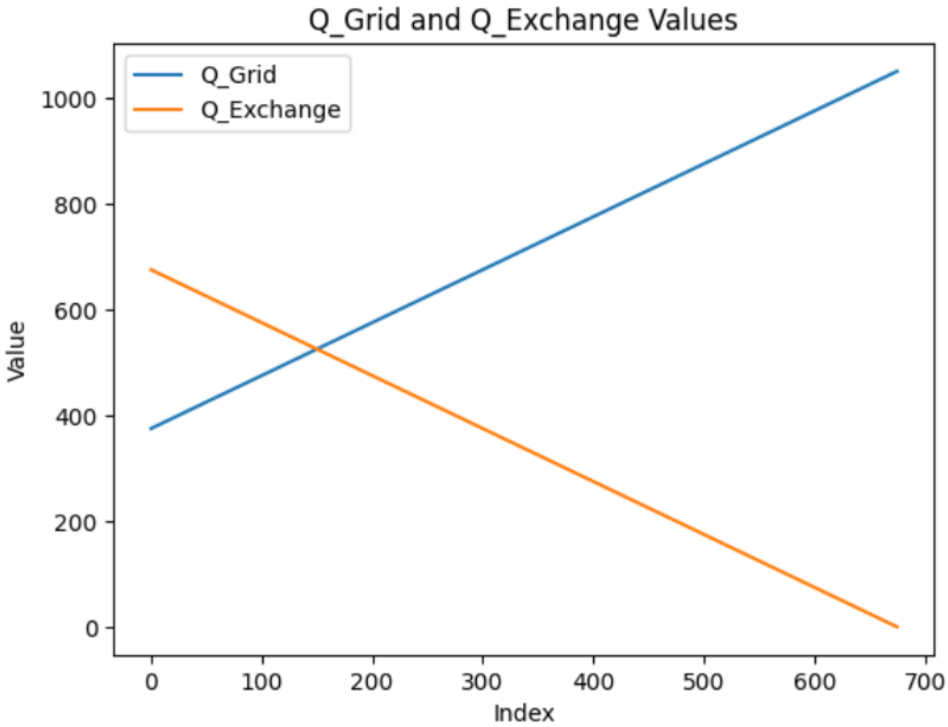
The analysis highlights the intricate relationship between the quantity of electricity drawn from the State Electricity Grid, the percentage of renewable energy, and the total price. Balancing these factors requires careful consideration and may involve trade-offs between sustainability objectives and economic constraints. Further optimization techniques and strategic planning are necessary to address this challenge effectively.

Methodology:

Two optimization methods were employed to determine the most efficient electricity usage strategy for January 1, 2018. The objective was to maximize the percentage of total renewable electricity while simultaneously minimizing the total cost.

Method 1:

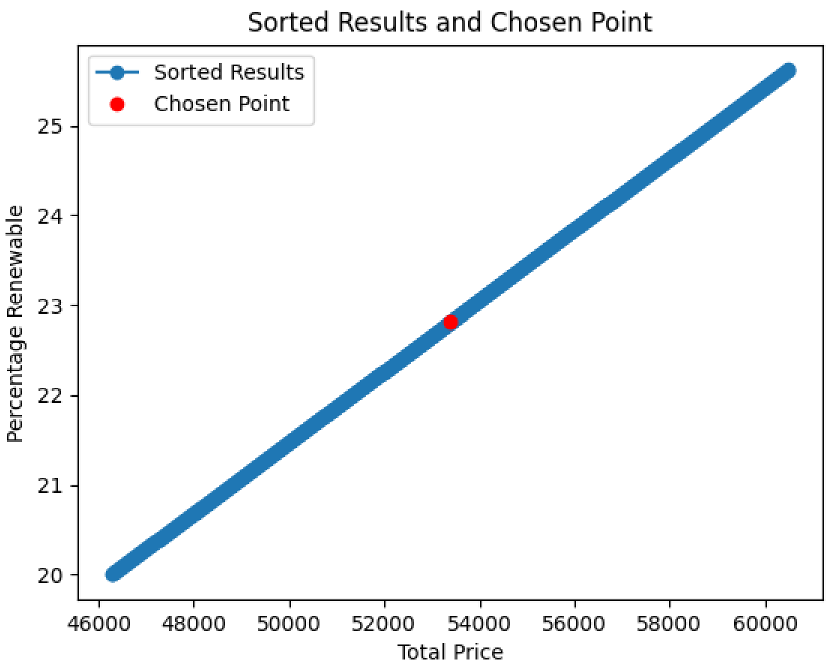
In the first approach, an exhaustive search was conducted to evaluate various combinations of quantities of electricity drawn from the State Electricity Grid (Q_Grid) and the Power Exchange (Q_Exchange). For each combination, the total price and the percentage of renewable electricity were calculated.



Points where the percentage of renewable electricity exceeded 20% were considered as potential solutions. Among these points, the intersection point of Q_Grid and Q_Exchange was identified. The total price at this intersection point was determined to be `49428.75`. The percentage of renewable electricity was `21.25`

Method 2:

The second method involved filtering the results from Method 1 to include only points where the percentage of renewable electricity was greater than or equal to 20%. These filtered results were then sorted based on the total price in ascending order.



To strike a balance between the percentage of renewable electricity and the total cost, a point closer to the middle of the sorted results was chosen. The chosen point had a percentage of renewable electricity of `22.816667` and a total price of `53393.67`.

Results:

a) Optimized Percentage of Total Renewable Electricity: `21.25 & 22.82`

b) Optimized Quantity of Electricity drawn from State Electricity Grid (Q_Grid): `525 & 713`

c) Optimized Quantity of Electricity drawn from the Power Exchange (Q_Exchange): `525 & 337`

Conclusion:

The optimized electricity usage strategy for January 1, 2018, balances the need for renewable energy with cost-effectiveness. By maximizing the percentage of renewable electricity while minimizing the total cost, this strategy contributes to sustainability goals while ensuring efficient resource utilization.