# Predictive Modeling for Public EV Charging Stations in Denmark

Emil Hu

Dept. of Electrical and Computer Engineering
Aarhus University
Arhus, Denmark
201907692@post.au.dk

Abstract—The increasing popularity of electric vehicles (EVs) has led to a significant increase in the demand for public charging stations. In this study, the load of the public charging infrastructure in Denmark is examined. This study employs three models, namely Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Bidirectional LSTM, to predict the load on the public EV charging stations. The models were trained and tested on data from various charging stations in Copenhagen and despite limitations in the data, all the models demonstrated satisfactory performance, particularly in forecasting the weekly charging load. Both the LSTM and BiLSTM model performed marginally better than the GRU model. The study's findings provide the groundwork for a forecasting model capable of accurately predicting the EV charging load with possibilities for future improvements.

Index Terms—electric vehicles, public charging stations, forecasting, recurrent neural network

## I. INTRODUCTION

The ongoing global climate crisis has proven to be a massive challenge requiring urgent need for sustainable solutions. One significant contributor to the environmental degradation is the transport sector. With the transport sector being a major contributor to greenhouse gas emissions, the transition towards electric vehicles (EVs) becomes a crucial factor mitigating these impacts. In Denmark, the interest and number of EVs have increased significantly in recent years. Recent numbers show that 42,6% of all newly registered cars in September was EVs, the highest percentage recorded to date [1]. However, with this rapid growth in EV popularity, the industry faces a 'chicken and egg' dilemma. This dilemma revolves around the need for sufficient public charging infrastructure to support the growing number of EVs, while at the same time, the expansion of such infrastructure is dependent on the assurance of users adopting EVs.

This article aims to explore and address this dilemma by developing and evaluating a multiple forecasting models for public EV charging station load in Denmark. This study is a part of a larger project aiming to predicting new locations for public charging stations. Thus, by accurately predicting the demand and usage patterns of these charging stations, this study seeks to inform and guide the strategic expansion of the public charging infrastructure. An accurate forecasting model also provides insights for optimizing the resource allocation,

optimize resource allocation, reduce operational costs, and enhance the overall user experience.

The rest of this paper is organized as follows: Section II provides an overview of existing research on EV charging load forecasting methods, highlighting key advancements in current methodologies. Section III and Section IV presents the forecasting models and the dataset used in this study, respectively. This also includes the necessary preprocessing steps undertaken to prepare the data for training. Section V details the training process of the forecasting models as well as the evaluation of the models. Lastly, Section VI discusses the obtained results, while VII provide concluding remarks on the project as well as potential future work.

## II. RELATED STUDIES

Several studies have conducted research on forecasting the load on charging stations. These studies employ various models and approaches to tackle this complex challenge. A study by Zhu et al. (2019) presents a comprehensive analysis of short-term load forecasting for a EV charging station using deep learning models. In the study, four deep leaning approached were implemented. This included Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU), and Deep Neural Networks (DNN), which were trained on data from a large EV charging station in Shenzhen, China, focusing on charging time, quantity, and real-time electricity prices. The models were evaluated using metrics like Normalized Root Mean Squared Error (NRMSE) and Normalized Mean Absolute Error (NMAE) and the study found that while all models showed limitations in predicting peak loads, the GRU model had comparatively better performance. The study suggests that incorporating more influencing factors will improve the prediction accuracy and speed of these models [2]. A different approach is presented in a study by Dabbaghiamanesh et al. (2021). This research proposes a forecasting model based on the Q-learning technique, a type of reinforcement learning, to predict the load demand of plug-in hybrid electric vehicles (PHEVs) under different charging scenarios. The study demonstrates that the O-learning technique can improve the accuracy and speed of load forecasting when compared to traditional techniques like ANN and RNN [3].

The findings from these studies underline the various possibilities in the field of EV charging station load forecasting. Common to both is the primary focus on short-term forecasting of the charging load. On the other hand, a study by Yi et al. (2022) explores long-term demand forecasting for EV charging stations. This study aims to forecast the demand for a 1-month horizon and a multi-step prediction for a 5-month horizon. The study employs a deep learning model, Sequence to Sequence (Seq2Seq), which achieves high prediction compared to traditional models like ARIMA and XGBoost [4]. The study's focus on monthly forecasting provides a valuable and different perspective to the short-term forecasting approaches.

While research focusing on forecasting the load of public charging station has not been conducted in Denmark, which is the geographical scope of this study, these studies provide valuable insights and knowledge in the field of EV load forecasting.

### III. FORECASTING MODELS

Predicting the load on public charging infrastructure is a complex task and multifaceted influenced by various factors such as the electricity price and seasonal variations. One common way to address this challenge is by implementing Recurrent Neural Networks (RNNs). Unlike traditional neural networks that process inputs independently and lack memory of previous inputs, RNNs contains feedback connections in the network architecture, which allows them to process sequences of data. In that way, RNNs are able to retain information over time, making them suitable for understanding and forecasting time-dependent data. For this project, two types of RNNs has been selected for evaluation: Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU). Additionally, Bidirectional LSTM (BiLSTM), which is an extension of LSTM, will also be evaluated.

# A. LSTM

LSTM was introduced by Hochreiter & Schmidhuber (1997) and was designed to handle the vanishing gradient problem, where gradients become exceedingly small during backpropagation. This makes it challenging for RNNs to learn from long sequences, also known as long-term dependencies. The LSTM model tackles by defining three gates in its network architecture:

- **Input gate** = Controls the extent of new information to be added to the cell state.
- Forget gate = Determines the information to be removed from the cell state.
- Output gate = Decides what information from the cell state should be used to generate the output at each time step.

This gating mechanism regulates the flow of information and allows for LSTMs to retain information over extended sequences and thus learn long-term dependencies [5].

BiLSTM extends the LSTM architecture by processing the data in both forward and backward directions. In the BiLSTM model, two separate LSTMs are applied: one that processes

the input sequence as it is, while the other processes a reversed copy of the input sequence. This configuration allows the model to capture information from both past and future contexts.

## B. GRU

The GRU model was introduced by Cho et. al (2014) and is similar to the LSTM model as it is also designed to address the vanishing gradient problem. However, instead of using a 'cell state' that regulates the information, hidden states are used. Additionally, instead of three gates, the network architecture comprises of the following two gates:

- **Reset gate** = Balances the information from the previous state and the new information from the current input.
- Update gate = Determines how much past information to forget.

Like LSTM, these gates control how much and what information that shall be retained. Although the architecture is more streamlined, resulting in a simpler model with fewer parameters [6].

#### IV. DATASET AND PREPROCESSING STEPS

For this project, Norlys A/S provided the dataset, which were gathered from ChargeFinder [7]. A data analysis of the dataset has previously been conducted by the author, analysing the charging pattern and current capacity of the public charging infrastructure in Denmark [8]. The dataset comprises of a parquet (parq) file and contains time-series data from observed charging stations. The dataset consists of 1705159 data points, which spans from August 10, 2023 to October 16, 2023 and have been sampled at 15-minute intervals. The dataset has the format:

[Slug, Stander, Timestamp, Owner, Hour, Date,

Where 'slug' is the ID of the charging stations and 'timestamp' is the time and date the data point was sampled. 'Hour', 'Date', and 'Day\_of\_week' are derived directly from the timestamp, representing the hour of the day, date, and day of the week the data was sampled, respectively. The 'owner' attribute describe the company that maintains the charging station and the 'longitude' and 'latitude' represent the geographical coordinates of the charging station. Finally, the 'Stander' parameter is a list of the charging points at the charging station. Each charging point is further described with a status code that describes the current state of the charging point, which can be either 0 (unknown), 2 (available), 3 (charging) or 5 (unavailable). Furthermore, each charging point either have an 'info' attribute or a 'price' attribute, which describe the duration a charging point has been at its current state and the current charging price, respectively [8]. Fig. 1 shows an example of two data samples, one with the 'info' attribute and one with the 'price' attribute.

```
wj6ex8,"[{'id': 'DK*CKE*E168*1', 'status': 2, 'price
     ': '4.50 DKK/kWh', 'tariffs': [{'name': 'Circle
    K', 'shortDescription': '', 'currency': 'DKK',
    planId': 'plpj5n6', 'costKwh': '4.50'}, {'name':
    'Circle K Drop-In', 'shortDescription': '', '
     currency': 'DKK', 'planId': 'plkjp86', 'costKwh
     ': 5.5}]}",2023-08-10 13:33:14.569419,13.0,
     Circle K, 2023-08-10, 13:33:14.569419, Thursday
     ,8.978481,56.108643
kp26k2,"[{'id': 'DKALLEGO0010161', 'status': 2, '
    price': 'Available: 1h 58m', 'free': 0, 'info':
    'Available: 1h 58m'}, {'id': 'DKALLEGO0010162', 'status': 2, 'price': 'Available: 31m', 'free': 0, 'info': 'Available: 31m'}, {'id': '
    DKALLEGO0010171', 'status': 2, 'price': 'Available: 30m', 'free': 0, 'info': 'Available:
     30m'}, {'id': 'DKALLEGO0010172', 'status': 2, '
    price': 'Available: 1h 50m', 'free': 0, 'info':
     'Available: 1h 50m'}]",2023-08-10
     13:33:35.204039,13.0,Allego
     ,2023-08-10,13:33:35.204039,Thursday
     ,10.124751,56.181601
```

Fig. 1: Data samples from the dataset. The upper data sample is from Circle K and displays charging station with the 'price' attribute. The lower data sample is from Allego and illustrates the 'info' attribute.

## A. Area Selection

While the dataset represents charging stations across the entire Denmark, only charging station within Copenhagen was selected for this study. This was due to the data analysis performed by the author being exclusively focused on Copenhagen as well as the high density of charging stations in the city, a characteristic only present in few parts of Denmark. Additionally, the dataset only spans two months, making it more suitable to analyze an area with a high concentration of charging stations to provide a better foundation for the forecasting model. By restricting the scope to charging stations in Copenhagen, the number of data points were decreased to 157144 data samples divided between 37 charging stations. Fig. 2 shows the charging stations located in the Copenhagen.

## B. Data Preprocessing Steps

In order to transform the raw data into a structured format, suitable for training the forecasting models, a number of data preprocessing steps were made. First, the primary variable of interest is the number of charging points with status code 3, representing the 'charging' status, at each charging station. This variable is determined by the ratio of charging points that were actively charging at the corresponding timestamp, providing a direct measure of charging station usage. The original dataset timestamps were in the format 'YYYY-MM-DD time\_of\_the\_day'. However, these timestamps were not usable for model training. Therefore, the time was represented by the hours elapsed since the first data sample. This transformation normalizes the time data, making it sequential and more suitable for time-series analysis. To achieve consistency in the time intervals, the data was divided into 30-minute



Fig. 2: Charging stations in the Copenhagen.

segments. Within each interval, the status code 3 (charging status) is averaged across all data samples that fall within that specific period. This approach ensures an even and consistent temporal resolution for the dataset. Furthermore, since the data was divided between 37 charging stations, it was necessary to group the data based on the charging station in order to forecast the load for each charging station. More specifically, the data was grouped based on the geographic coordinates of each charging station. As mentioned, some charging stations included the 'info' attribute, while others reported 'price' data. However, since these attributes were not uniformly present across all data points, they were excluded from the analysis to maintain data consistency and focus on the most relevant variables. Other attributes that were disregarded included the 'Hour', 'Slug', 'Owner', and 'Date' columns. Finally, it was decided to incorporate the day of the week into the model, which was done by one-hot encoding the categorical day data.

## C. Auto Correlation

A significant part of the aforementioned data analysis conducted by the author was determining charging patterns. Fig.3 showcases a histogram of the daily status counts for charging stations, which is a part of the data analysis. The graph shows the counts for the 'charging' and 'available' statuses and the numbers on top represent the percentage of 'Charging' statuses out of the total number of statuses recorded at each hour. The numbers show that more charging stations are in use between the hours 10-15, while this activity decreases as the day progresses [8]. This pattern suggests that there are daily and potential weekly patterns in the user charging behaviour.

Therefore, to check if these temporal patterns happened regularly, an autocorrelation analysis was conducted. Autocorrelation analysis is a statistical method used to determine the degree to which current values in a time series are related to past values at successive time intervals, known as lags.

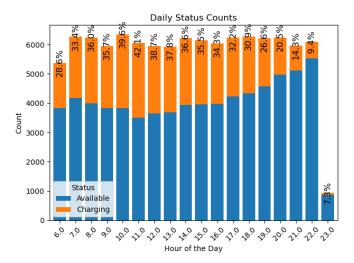


Fig. 3: Daily status count for charging stations in Copenhagen [8].

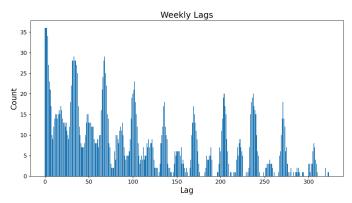


Fig. 4: Count of significant lags for every charging station in Copenhagen.

Autocorrelation can reveal underlying patterns in the data such as trends or seasonal cycles.

Fig. 4 presents the results of the autocorrelation analysis, displaying the count of each significant lag across all charging stations. Each bar on the histogram corresponds to the count of significant autocorrelation instances at various lags, with each lag representing a 30-minute interval. As seen on the graph, there are noticeable peaks at regular intervals, which are indicative of recurring usage patterns. Furthermore, there is a peak at lag 336, which corresponds to one week, and thereby confirms a weekly pattern. The findings from the analysis point to regular usage patterns that could be driven by daily routines and weekly schedules. Based on these results, the significant lags were incorporated as features into the forecasting models. Fig. 5 shows the exact lags that were incorporated.

# D. Sequence Creation

The final preprocessing step prior to model training involved the creation of input sequences tailored for the forecasting models. The process of sequence creation is designed to significant\_lags = [0, 1, 2, 32, 66, 100, 144, 170, 205, 236, 288, 336]

Fig. 5: Significant lags

structure the time-series data into a format that these models can effectively learn from. As previously mentioned, the data was grouped based on the geographical coordinates of each charging station. Likewise, the sequences were generated by iterating over the grouped data such that each sequence only contained data from a single charging station. This approach allows the models to capture temporal patterns specific to each location. The sequence length was set to 336 intervals, equivalent to one week of data, capturing the weekly cycles as identified in the autocorrelation analysis. This length was chosen to allow the models to learn from weekly patterns. In total, the sequence creation process yielded 56,253 sequences.

## V. MODEL TRAINING AND EVALUATION

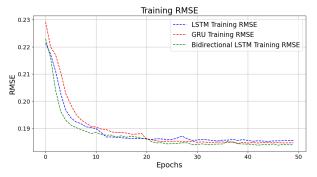
The aim of the forecasting models is to predict the load i.e. the ratio of charging points that are charging for each of the charging stations. As stated, the models considered for this study is LSTM, BiLSTM and GRU.

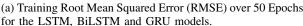
## A. Training Strategy and Model Configuration

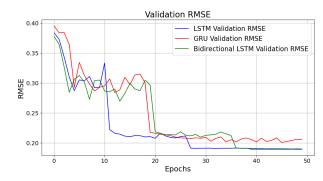
To train the models, the dataset has been split into three sets: 80% for training, 10% for validation and 10% for testing. Since the dataset consists of time-series data, the dataset was divided into contiguous blocks to maintain the temporal order of the data. To ensure a consistent comparison between the models, each model is structured with a layered architecture consisting of two hidden layers of size 50 and dropout rate set to 0.5 to prevent overfitting. RMSPROP is the chosen optimizer to adjust weights efficiently, with a learning rate of 0.0001. The learning rate is complemented with a scheduler, which adjusts the learning rate based on validation loss and improve convergence during training. Furthermore, each model has been initialized with Xavier uniform for weights and uniform initialization for biases. Lastly, the data has been processed with a batch size of 32 and the models have been trained over 50 epochs. This setup allows for an objective evaluation of each model's ability to forecast the load on the charging stations without the influence of varying hyperparameters.

### B. Model Performance

In assessing the performance of the LSTM, GRU, and Bidirectional LSTM models, the Mean Squared Error (MSE), Mean absolute error (MAE) and Root Mean Squared Error (RMSE) are used as accuracy metrics. The MAE is the average of the absolute difference between the true and predicted values. It measures the average of the residuals in the dataset. MSE is the average of the squared difference between the true and predicted values and measures the variance of the residuals. RMSE is the square root of the MSE and measures the standard deviation of the residuals [9].







(b) Validation Root Mean Squared Error (RMSE) over 50 Epochs for the LSTM, BiLSTM and GRU models.

Fig. 6: RMSE values for training and validation.

Model	MSE	RMSE	MAE
LSTM	0.0363	0.1908	0.1417
GRU	0.0410	0.2026	0.1621
BiLSTM	0.0366	0.1913	0.1453

TABLE I: Output accuracy metrics for each model after testing.

Fig. 6a shows the RMSE values during training. As seen on the graph, all models showed a quick decline in RMSE with BiLSTM converging marginally faster than the other two models. However, the difference between each model seems to be minimal. This indicate that the models are able to capture the temporal patterns early in the training process. Likewise, Fig. 6b shows the model RMSE values when tested on the validation set. Initially, each model experiences fluctuations in RMSE. However, after a few epochs there is a marked decline, with LSTM model showing a steeper decrease. As the epochs advance, the validation RMSE of all models converges. Both LSTM and BiLSTM stabilizes at a lower RMSE compared to GRU, suggesting that additional complexity of the LSTM models may provide a slight advantage in capturing the patterns in the data.

Upon completion of the training phase, the models were evaluated on their performance on the test set. Tab. I shows the MSE, RMSE and MAE for each model. As seen, the LSTM model achieved an MSE of 0.0363, RMSE of 0.1908, and MAE of 0.1417, indicating a strong predictive ability with relatively low error magnitude. The BiLSTM obtained similar values, though with a marginal increase in error compared to the unidirectional LSTM model. Lastly, the GRU model, with an MSE of 0.0410, RMSE of 0.2026, and MAE of 0.1621, showed slightly higher errors, which may be due to its more simple architecture compared to the other two models. In general, these metrics provide a comprehensive assessment of each model's performance. While the LSTM model exhibited the lowest errors, the difference between the models were minimal. Furthermore, the close performance of the LSTM and BiLSTM models is also noteworthy as it suggests that the

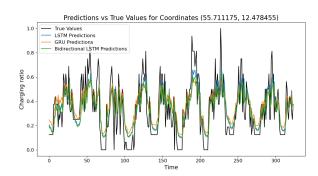


Fig. 7: Testing set model predictions for charging station with coordinates (55.711175, 12.478455).

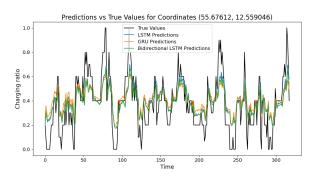


Fig. 8: Testing set model predictions for charging station with coordinates (55.67612, 12.559046).

additional complexity of the BiLSTM does not translate to a significant improvement for this forecasting task.

The final evaluation of the forecasting models involved analyzing their predictions against the true values from the test set. The test set comprised of approximately one week's worth of data. Fig. 7 and Fig. 8 display the predictions for two different coordinate groups, which represent individual charging stations. In both figures, all the models are able to follow the overall pattern with high accuracy, but have trouble reaching the high peaks. Overall, the LSTM and BiLSTM models seems to perform equally good, while the GRU model

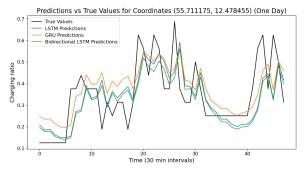


Fig. 9: 24 hour model predictions for charging station with coordinates (55.67612, 12.559046).

deviates slighty more compared to the other two models. It is especially around the declines, where the GRU model has trouble predicting the charging load.

In addition to evaluating model performance over an entire week, an analysis of the predictions for single day was also conducted in order to understand the models' precision on a daily basis. Fig. 9 illustrates the predictions by each model along with the actual values over single day for a specific charging station. Like the findings shown in Fig. 7 and Fig. 8, the models are able to capture the overall trend of the charging station load. Again, the LSTM and BiLSTM model perform nearly identically, while the GRU model deviates more. However, for the single day predictions, the gap between models' predictions and the actual values are more pronounced. Furthermore, it also appears that there is a slight lag in the models' response to sudden increase or decrease in the charging ratio.

These graphs serve as a visual confirmation of the model's performance and suggest that all models can forecast the charging station load with a high degree of accuracy, with a slight edge to LSTM and BiLSTM models. However, the added complexity of the BiLSTM model does not seem provide a significant improvement on the overall forecasting accuracy.

## VI. DISCUSSION

The evaluation of the forecasting models has revealed several insights and limitations, providing a key understanding of their current performance.

Overall, the LSTM, GRU, and BiLSTM models demonstrated satisfactory performance, achieving reasonable accuracy in forecasting the load on charging stations in the city of Copenhagen. The models obtained low values of MAE, RMSE, and MSE, which was further supported by the visual inspection of the models' predictions against the test set values, where they were capable of capturing the overall patterns. However, a notable observation was the models' inability to reach the highest peaks of demand, which suggests that their ability to handle sharp, short-term fluctuations could be improved. Additionally, the models were better at capturing and predicting weekly patterns as opposed to daily patterns,

which shows that learning long-term dependencies is the primary strength of the models.

While the models are generally robust, they do posses some limitations. This include both in the dataset and the models itself. The dataset contained some inconsistencies in the type of information recorded, where some data points included the 'info' attribute, while others included the 'price' data. Furthermore, with the dataset only containing two month of available data, this resulted in the models were only fed significant lag times and day-of-week data as features, while potentially valuable features were disregarded. The absence of detailed feature engineering, such as including price and charging duration, likely impacted the models' ability to capture daily patterns more accurately. The lack of data also meant the models had limited ability to learn and generalize from seasonal trends, which may occur over longer time horizons. Another simplification was the preprocessing of the data into 30-minute intervals and setting the average charging ratio within that interval as the target variable. These steps could potentially oversimplify the complexity of the charging station use and result in the loss of nuanced information for the models' predictions. Lastly, due to the absence of live data, real-time forecasting was not feasible. Instead, the study relied on the test set to evaluate model performance, which may not fully represent the models' capabilities in an operational setting.

The findings from this study demonstrate the current capabilities of the implemented forecasting models, but do also highlight several possibilities for future research and development improvements. One key area is the exploration of alternative models such as Prophet or Seasonal Autoregressive Integrated Moving Average (SARIMA), which are specifically designed for time-series data and may provide enhanced capability in capturing seasonal patterns. Alongside this, a more data-driven strategy to hyperparameter tuning should be adopted in order to enhance the accuracy and efficiency of the models. Likewise, experimenting with different optimization algorithms such as ADAM, could yield potential improvements over the current RMSPROP optimizer. Another crucial improvement is collecting more data. This would not only improve the accuracy of the models, but also allow for incorporating more features such as 'info' and 'price. It would also allow for extending the geographical scope to include the entirety of Denmark. Additionally, integrating external factors, such as the increasing number of electric vehicles on the road and the number of fast chargers at each station, could provide a more comprehensive understanding for the models as these factors could have a significant impact on the charging station usage.

Thus, while the models do show promise, there is possibilities to build upon the current foundation laid by the current study by addressing the existing limitation and following the proposed future work.

## VII. CONCLUSION

The purpose of this study was to implement forecasting models capable of predicting the load of public charging stations in the city of Copenhagen, Denmark. This study employ three machine learning models, Long Short-Term Memory (LSTM), Bidirectional Long Short-Term Memory (BiLSTM), and Gated Recurrent Units (GRU). All models obtained low Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) and were able to capture the overall patterns when comparing their predictions to the test set values. Overall, all the models demonstrated satisfactory performance, where LSTM and BiLSTM performed marginally better than GRU. However, all models had trouble reaching predicting the high peaks and the models also performed better on weekly load predictions as opposed to daily predictions. These problems could be addressed by collecting more comprehensive data. The inconsistency and limited scope of the available data meant that valuable features could not be fully integrated in the models, which could further enhance the accuracy and effectiveness of the models.

#### REFERENCES

- [1] Danmarks Statistik, "NYT: Bestanden af elbiler steg med 69 pct.
  i 2022," December 2022. [Online]. Available: https://www.dst.dk/da/Statistik/nyheder-analyser-publ/nyt/NytHtml?cid=40083
- [2] J. Zhu, Z. Yang, Y. Guo, J. Zhang, and H. Yang, "Short-Term Load Fore-casting for Electric Vehicle Charging Stations Based on Deep Learning Approaches," *Applied Sciences*, vol. 9, no. 9, p. 1723, Jan. 2019.
- [3] M. Dabbaghjamanesh, A. Moeini, and A. Kavousi-Fard, "Reinforcement Learning-Based Load Forecasting of Electric Vehicle Charging Station Using Q-Learning Technique," *IEEE Transactions on Industrial Infor*matics, vol. 17, no. 6, pp. 4229–4237, Jun. 2021.
- [4] Z. Yi, X. C. Liu, R. Wei, X. Chen, and J. Dai, "Electric vehicle charging demand forecasting using deep learning model," *Journal of Intelligent Transportation Systems*, vol. 26, no. 6, pp. 690–703, Oct. 2022.
- [5] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 11 1997. [Online]. Available: https://doi.org/10.1162/neco.1997.9.8.1735
- [6] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling," 12 2014.
- [7] "ChargeFinder Ladestandere til elbiler." [Online]. Available: https://chargefinder.com/dk
- [8] E. Hu, "Data Analysis of Public Charging Infrastructure and User Behaviour in Denmark," 11 2023.
- [9] A. Chugh, "MAE, MSE, RMSE, Coefficient of Determination, Adjusted R Squared — Which Metric is Better?" Mar. 2022.