

Northeastern University, Boston, MA

**XGBoost-Driven Demand Prediction for Optimized Electric
Vehicle Charging Recommendations**

A project submitted in partial fulfilment of the requirements for the course
CIVE 7381 (Transportation Demand Forecasting and Model Estimation)

Nathan David Obeng-Amoako, Zhaoming Zeng, Qi Kang

December 2024

Table of Contents

1	Introduction	1
2	Dataset	4
2.1	Dataset Overview	4
2.2	Timeframe and Records	4
2.3	Charging Duration Analysis	5
2.4	Feature Engineering	6
2.5	Data Aggregation	7
2.6	Charging Session Counts per Station	7
3	Prediction Model	9
3.1	XGBoost Model	9
3.2	Feature Selection	11
3.3	Prediction Performance Comparison	14
3.3.1	Comparison Between OLC Regression and XGBoost	14
3.3.2	Comparison among All Feature Sets	18

3.3.3	Generalization Test	19
3.4	Summary	21
4	EV Charging Station Recommendation System	23
4.1	Problem Statement	23
4.1.1	Constraints	24
4.1.2	Objective Function	26
4.1.3	Mathematical Formulation	27
4.2	Case Study and Numerical Result Analysis	28
5	Conclusion	30

List of Figures

1	Distribution of Charging Session Durations for JPL and Caltech	6
2	Charging Session Distribution Across Stations at JPL	8
3	Charging Session Distribution Across Stations at Caltech	8
4	Converged Learning Curve for XGBoost Model	15

5	Regression vs XGBoost	15
6	Importance Ranking for XGBoost Model	17
7	Feature Sets Comparison	18
8	Generalization	20
9	Feature Importance Ranking Comparison	21
10	Introduction of the EV Charging Station Recommendation System	24
11	EV Charging Station Price Rates over a 24-hour Period	28
12	Results of Recommendation System	29

List of Tables

1	Field Descriptions for the ACN Dataset	5
2	Feature Selection	12
3	Regression T-statistics Check(bold for $ t > 2$)	16
4	Parameter Value for MIP Formulation	29

1 Introduction

The increasing adoption of electric vehicles (EVs) has been a significant development in the push toward sustainable transportation. However, the rapid rise in EV ownership has placed considerable strain on the existing charging infrastructure, highlighting the urgent need for effective solutions to optimize the utilization of charging stations. This shift introduces new challenges: the limited availability of charging points, increased waiting times, and fluctuating energy demands at charging stations. Additionally, the advent of dynamic pricing, where the cost of charging increases with demand, complicates the decision-making process for users, who must balance cost, charging speed, and station availability when selecting a location and duration for charging. As the number of EVs continues to grow, it becomes critical to implement intelligent systems that optimize the charging process in real-time, enhancing user satisfaction while reducing congestion and improving infrastructure efficiency.

The management of EV charging infrastructure has emerged as a complex problem that demands the integration of advanced technologies such as machine learning, optimization algorithms, and real-time data processing. A fundamental aspect of managing this infrastructure is demand forecasting, where the goal is to predict future occupancy levels at charging stations to inform decision-making. Accurate demand predictions are crucial, as they

allow for proactive management of resources, minimizing long wait times and optimizing charging durations. Furthermore, real-time personalized recommendations for EV users are necessary to guide them toward optimal charging decisions, ensuring cost-effectiveness and time efficiency while considering user-specific needs (e.g., battery levels, preferred charging duration, and location).

In the context of these challenges, this project seeks to develop an integrated framework that combines predictive modeling with optimization to provide EV users with actionable insights for real-time decision-making. The Adaptive Charging Network (ACN) dataset from the California Institute of Technology (Caltech), which includes extensive charging session logs, timestamps, energy deliveries, and site-specific identifiers, serves as the main source of data for this project. By utilizing this dataset, we aim to build a forecasting model based on Extreme Gradient Boosting (XGBoost) to predict future charger occupancy at key stations and then feed these predictions into an optimization model that recommends personalized charging strategies. These strategies will take into account dynamic pricing, the capacity of each station, predicted demand, and user-specific constraints, such as remaining battery charge and geographic location.

This work builds upon existing research in the areas of demand forecasting and optimization but goes further by integrating these methods into a holistic real-time recommender system that adapts to the dynamic nature of EV

charging environments. Although previous studies have examined various aspects of EV charging, such as station selection and cost optimization, few have effectively incorporated dynamic pricing models into real-time optimization for personalized decision-making. Furthermore, while demand prediction using machine learning techniques like XGBoost has shown promise in other domains, its application in EV charging infrastructure management has not been explored to the same extent.

The primary goal of this research is to contribute to the scalable management of EV charging infrastructure, providing both EV users and charging station operators with tools to enhance the efficiency of the charging process. Specifically, we aim to reduce user costs, wait times, and overcrowding at stations, improving the overall user experience and contributing to the sustainability of electric transportation. In doing so, this work lays the foundation for future advancements in smart charging systems, which are critical to the long-term success of EV adoption and the broader EV ecosystem.

2 Dataset

2.1 Dataset Overview

The ACN dataset ¹, as detailed by Lee, Li, and Low 2019, offers comprehensive information on EV charging sessions, enabling studies into efficient EV charging algorithms and infrastructure optimization. Based on a deployment of over 100 charging stations at 3 sites around the Caltech campus, the ACN system serves an average of 65 EVs per day under a shared power constraint of 300 kWh. Between 2017 and 2019, it supported the delivery of energy sufficient for 2.3 million miles of charging. For this project, we only use data from two sites; viz., Caltech, which has 53 Electric Vehicle Supply Equipment (EVSEs) and JPL, which has 52 EVSEs. These datasets contain metadata and nested structures, capturing various attributes relevant to the study. The JSON files were normalized and transformed into pandas DataFrames for structured analysis. The relevant fields included in the dataset are presented in **Table 1**.

2.2 Timeframe and Records

We utilized three months of data spanning January to March 2019. The first two months were designated for training, and the final month was used for

¹<https://ev.caltech.edu/dataset>

Field	Description
<code>_id</code>	Unique identifier of the session record
<code>connectionTime</code>	Time when the EV plugged in
<code>disconnectTime</code>	Time when the EV plugged out
<code>doneChargingTime</code>	Time when of the last non-zero current draw recorded
<code>sessionID</code>	Unique identifier for the session
<code>siteID</code>	Unique identifier for the site
<code>stationID</code>	Unique identifier of the EVSE
<code>kWhRequested</code>	Energy requested by the user in kWh
<code>kWhDelivered</code>	Amount of energy delivered during the session
<code>timezone</code>	Timezone of the site. Based on pytz format.

Table 1: Field Descriptions for the ACN Dataset

testing. The JPL dataset contains 4,117 session records, while the Caltech dataset comprises 2,907 session records. Missing values accounted for less than 0.5% of the data and were removed to maintain data integrity.

2.3 Charging Duration Analysis

The average charging duration at JPL was calculated to be 7.34 hours, whereas the corresponding average at Caltech was 5.83 hours. **Figure 1** illustrates a combined histogram showing the distribution of charging session durations for both locations. The histogram reveals that JPL sessions generally tend to be longer, possibly due to differences in user behavior or station utilization policies.

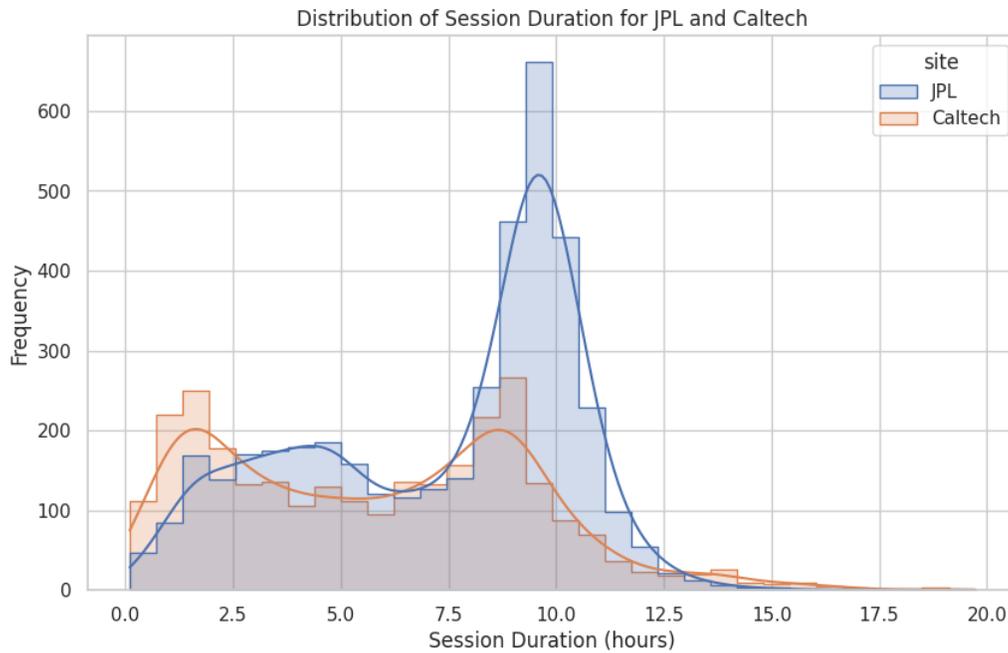


Figure 1: Distribution of Charging Session Durations for JPL and Caltech

2.4 Feature Engineering

Several new features were derived to enhance the predictive capacity of the dataset:

- **Day of the Week, Day of the Month, Month of the Year:** Derived from the connection and disconnection timestamps to capture temporal patterns.
- **Charging Duration:** Computed as the difference between connection and disconnection times.

2.5 Data Aggregation

The raw data was aggregated into half-hourly intervals to facilitate temporal analysis and model compatibility. This aggregation involved summarizing session counts and durations within each time block.

2.6 Charging Session Counts per Station

We analyzed the number of charging sessions at each station within JPL and Caltech, revealing distinct usage patterns. **Figures 2** and **3** display the distribution of session counts across approximately 50 stations at each location. At JPL, the distribution of charging sessions is nearly uniform across the stations, suggesting that all stations are utilized at similar rates. This uniformity may be attributed to high demand evenly distributed across the available infrastructure. In contrast, Caltech exhibits significant variability in station usage, with certain stations experiencing higher activity. This disparity could result from user preferences, station proximity to key locations, or varying accessibility.

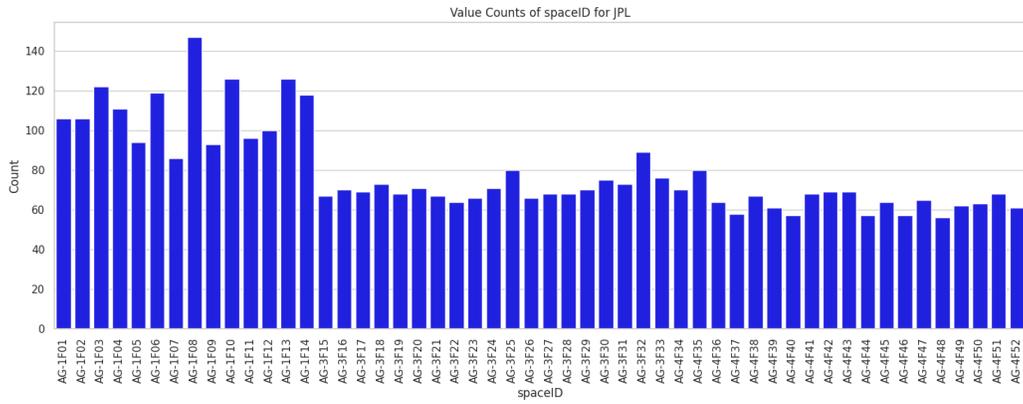


Figure 2: Charging Session Distribution Across Stations at JPL

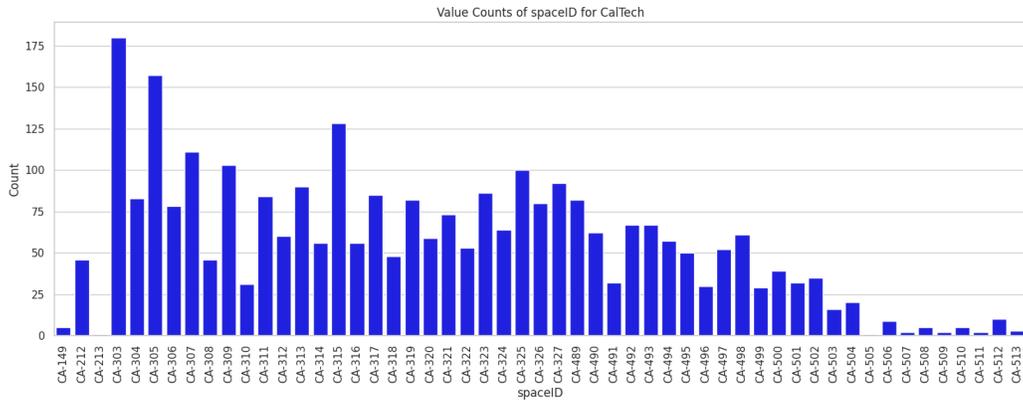


Figure 3: Charging Session Distribution Across Stations at Caltech

3 Prediction Model

In this section, we will introduce the methodology of the XGBoost model and use it as our prediction module. The feature selection is discussed in Section 4.2, and a Shapley value-based method is deployed to rank the feature's importance. The experiment results are analyzed in Section 4.3. Section 4.4 summarizes the investigation of the prediction model.

3.1 XGBoost Model

After confirming the prediction task, we need to find an efficient model to map the inputs to future charging demand. We can formulate the problem into a supervised learning task and adopt the popular machine learning algorithm extreme gradient boosting(XGBoost) as our solution. XGBoost stands out in supervised learning due to its superior performance and accuracy, stemming from its robust gradient-boosting framework and regularization features that combat overfitting. It offers remarkable efficiency and scalability, handling large datasets with speed across multiple processing units. Its flexibility in managing different data types, handling missing values, and allowing for customized objective functions makes it versatile for a wide range of applications. Moreover, XGBoost's built-in model interpretation tools aid in understanding feature influences, making it a preferred choice among data scientists for achieving top-notch predictive models in competitions and

real-world tasks.

Similar to other gradient-boosting algorithms, XGBoost relies on the decision tree ensemble structure. A tree ensemble model uses multiple decision trees to improve accuracy, as a single tree typically falls short. Each tree contributes to the final prediction by assigning scores to its leaves, which are then combined to make a more reliable prediction. With K number of decision trees f_k to predict the outcome \hat{y}_i from instance x_i , the model can be written as, where \mathcal{F} is a set of all possible trees.

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in \mathcal{F} \quad (1)$$

Then we can train the model by properly defining the loss function and optimizing it. The optimization object J can be set by taking both loss L from ground truth y_i and regularization ω :

$$J = \sum_{i=1}^n L(y_i, \hat{y}_i^{(K)}) + \sum_{i=1}^K \omega(f_i) \quad (2)$$

Since it is intractable to learn all the trees at once, we use an additive strategy by adding a new tree at each time step t :

$$\hat{y}_i^{(t)} = \sum_{k=1}^t f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i) \quad (3)$$

The new tree can be selected by optimizing the object defined below.

$$J = \sum_{i=1}^n \left[L(y_i, \hat{y}_i^{(t-1)}) + f_t(x_i) \right] + \omega(f_t) + \text{constant} \quad (4)$$

For general loss functions, we can take the Taylor expansion of the original equation and remove the constant parts

$$J = \sum_{i=1}^n \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \omega(f_t) \quad (5)$$

where g_i, h_i and $\omega(f_t)$ are defined below, T is the number of leaves and ω is the score on each leaf, γ, λ are control factors:

$$g_i = \frac{\partial L(y_i, \hat{y}_i^{(t-1)})}{\partial \hat{y}_i^{(t-1)}} \quad (6)$$

$$h_i = \frac{\partial^2 L(y_i, \hat{y}_i^{(t-1)})}{\partial (\hat{y}_i^{(t-1)})^2} \quad (7)$$

$$\omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (8)$$

3.2 Feature Selection

The feature selection is initially based on intuition, and we developed four sets of features. Their performance is evaluated based on rooted mean squared error(RMSE), and the models are also examined by their feature importance ranking calculated by the Shapley-Value-based method. The feature selection

for each set is shown below:

ID	Freq.	History	Hour of Day	Day of Week	Day of Month	Month of Year
1	60 min	12	✓	✓	✓	✓
2	30 min	12	✓	✓	✓	✓
3	30 min	24	✓	✓	✓	✗
4	30 min	24	✓	✓	✓	✗

Table 2: Feature Selection

The **Freq.** means the resolution of the date and the **History** represent the number of previous demands we use for the next time-unit prediction. The swoosh in the categorical feature means it is selected, and the crossing means it is not.

To explain and gain more insights into the model, we will use Shapley Value based method to rank the feature importance after getting the fitted model.

Shapley Value-based approach formulates the prediction model as a coalition game and features value as players. It's normally used for analyzing a single instance. It draws the interpretation from the payout of each feature, namely Shapley values. The overall Shapley value of an instance is defined as the difference between the prediction and average predictions of the whole dataset. The estimation process for the Shapley value of a feature of interest is demonstrated in Algorithm 1

After getting the Shapley values of each feature in every instance, we can use them to evaluate the feature's importance of the prediction model and the

Algorithm 1: Shapley Value Estimation

Input: Dataset of interest X ; XGBoost model \hat{f} ; Instance of interest x ; Feature index j ; Number of iterations M ; Number of features m ;

Output: Shapley value of the j -th feature $\phi_j(x)$

```
1 for  $i \in \{1, \dots, M\}$  do
2   Draw random instance  $z$  from  $X$ ;
   Permute the instance by random order:
    $x_o = (x_{(1)}, \dots, x_{(j)}, \dots, x_{(m)})$ ;
    $z_o = (z_{(1)}, \dots, z_{(j)}, \dots, z_{(m)})$ ;
   Construct two new instances:
    $x_{+j} = (x_{(1)}, \dots, x_{(j-1)}, x_{(j)}, z_{(j+1)}, \dots, z_{(m)})$ ;
    $x_{-j} = (x_{(1)}, \dots, x_{(j-1)}, z_{(j)}, z_{(j+1)}, \dots, z_{(m)})$ ;
    $\phi_j^i = \hat{f}(x_{+j}) - \hat{f}(x_{-j})$ ;
3  $\phi_j(x) = \frac{1}{M} \sum_{m=1}^M \phi_j^m$ 
```

particle dependence on the feature of interest.

The idea behind feature importance evaluation is simple: Features with large absolute Shapley values are important. Since we want the global importance of the feature I_j , we average the absolute Shapley values per feature across the whole dataset:

$$I_j = \frac{1}{n} \sum_{i=1}^n \left| \phi_j^{(i)} \right| \quad (9)$$

Then, we can sort the features by decreasing importance and plot them, which provides a clear way to understand which feature plays a vital role in the prediction, in other words, influences the driver's decision most.

3.3 Prediction Performance Comparison

In this section, we will conduct three pairs of comparisons. The first one is the prediction comparison between the OLC regression model and the XGBoost model using feature set 1. Then, we conducted a comparison among the different sets of features to find the most efficient feature set. Lastly, we did a generalization test on the best model we trained from last experiment. The test is conducted using another data source.

3.3.1 Comparison Between OLC Regression and XGBoost

In this part, we will compare the performance of the regression model and XGBoost model using RMSE on the test set. The data we use to compare prediction accuracy is the hourly charging occupancy from Jan 1, 2019, to March 30, 2019, at the JPL charging site. The data is split into the training set and the test set by a ratio of 0.8:0.2. The feature we input to the model is feature set 1, and categorical features are treated as dummy variables. The converged learning curve of the XGBoost Model is shown below:

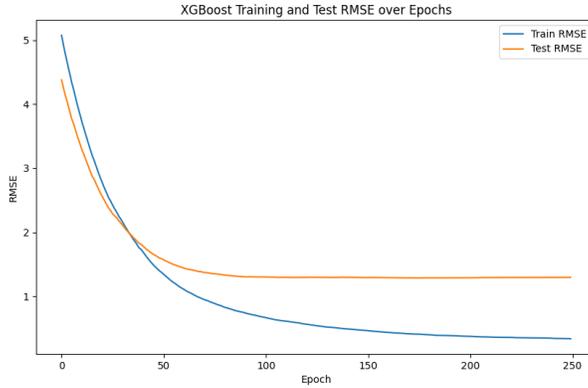
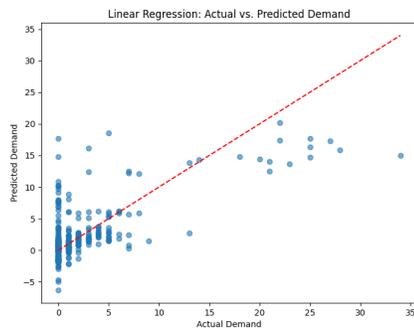
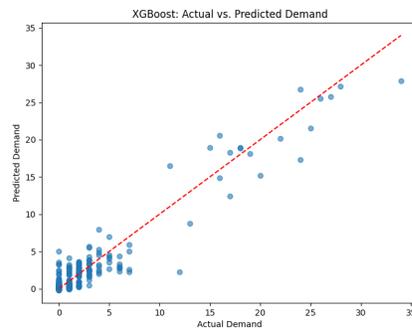


Figure 4: Converged Learning Curve for XGBoost Model

The scatter plots of both models are shown below; the caption of the plot is their RMSE value.



(a) Regression Model: 1.82



(b) XGBoost Model: 1.29

Figure 5: Regression vs XGBoost

As we can see from the plots, the points are distributed evenly alongside the 45-degree slopes, and it achieves better accuracy. The t-statistics for the regression model is shown below:

Feature Name	t-statistics	Feature Name	t-statistics
const	1.698	lag_1	14.857
lag_2	-5.908	lag_3	-2.184
lag_4	-0.643	lag_5	0.154
lag_6	-2.296	lag_7	0.115
lag_8	0.253	lag_9	-0.191
lag_10	-0.656	lag_11	-0.643
lag_12	-0.988	hour_of_day_1	5.296
hour_of_day_2	-0.531	hour_of_day_3	-0.677
hour_of_day_4	-1.727	hour_of_day_5	-1.653
hour_of_day_6	-1.769	hour_of_day_7	-1.847
hour_of_day_8	-1.882	hour_of_day_9	-1.926
hour_of_day_10	-1.852	hour_of_day_11	-1.810
hour_of_day_12	-0.988	hour_of_day_13	5.296
hour_of_day_14	14.516	hour_of_day_15	10.858
hour_of_day_16	-2.921	hour_of_day_17	-2.684
hour_of_day_18	0.430	hour_of_day_19	-0.697
hour_of_day_20	0.711	hour_of_day_21	0.593
hour_of_day_22	0.186	hour_of_day_23	0.332
hour_of_day_24	0.581	day_of_week_1	1.076
day_of_week_2	1.424	day_of_week_3	-0.274
day_of_week_4	0.466	day_of_month_5	0.131
day_of_month_6	0.217	day_of_month_7	1.281
day_of_month_8	0.346	day_of_month_9	-0.094
day_of_month_10	1.392	day_of_month_11	0.315
day_of_month_12	-0.057	day_of_month_13	-0.123
day_of_month_14	-0.146	day_of_month_15	-0.005
day_of_month_16	0.022	day_of_month_17	-0.695
day_of_month_18	0.677	day_of_month_19	0.185
day_of_month_20	-0.910	day_of_month_21	1.189
day_of_month_22	-0.062	day_of_month_23	0.138
day_of_month_24	1.130	day_of_month_25	0.498
day_of_month_26	0.532	day_of_month_27	0.225
day_of_month_28	0.767	day_of_month_29	-0.041
day_of_month_30	0.952	day_of_month_31	0.306
month_of_year_2	0.387	month_of_year_3	0.387

Table 3: Regression T-statistics Check(bold for $|t| > 2$)

For the regression model, all of the date-related features are trivial, and only the history demand and hour-of-day are significant. The feature ranking for the XGBoost model is shown below:

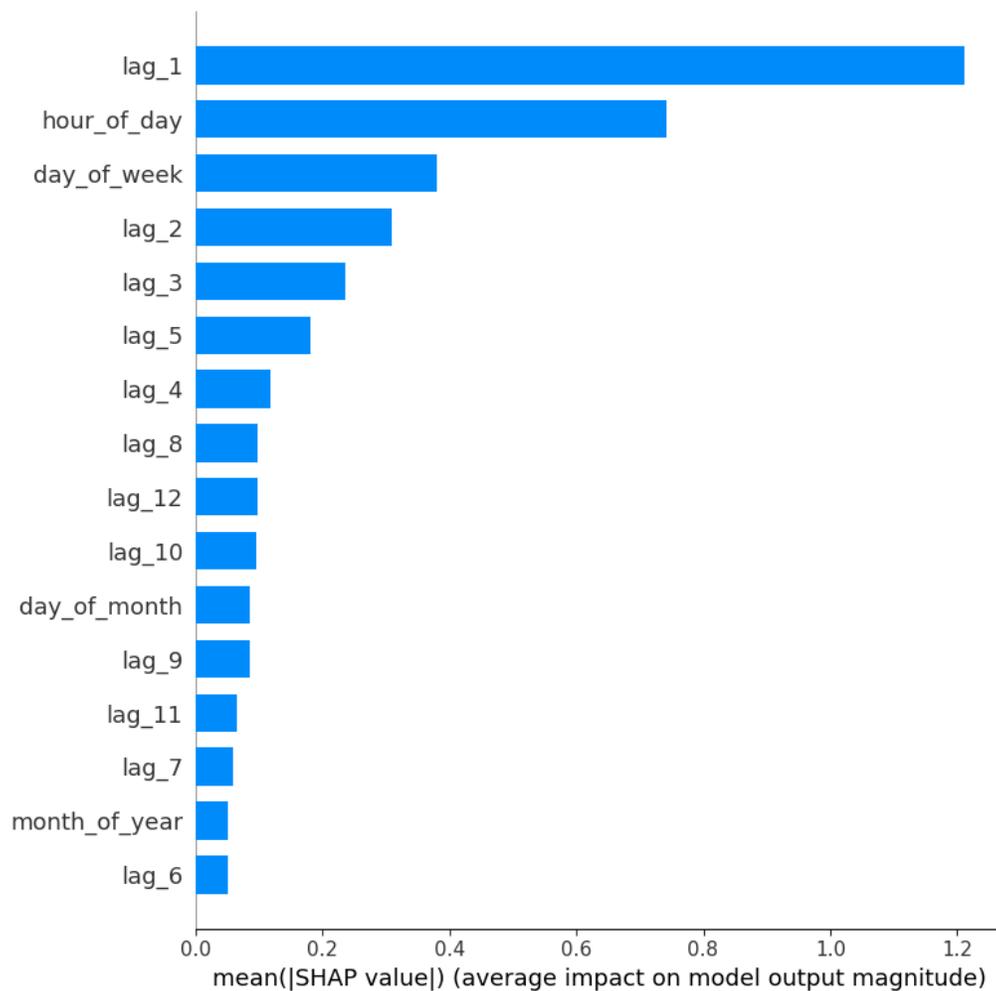


Figure 6: Importance Ranking for XGBoost Model

The XGBoost Model's top three features are the demand in the previous

hour, the hour of the day, and the day of the week. It considers more the day-of-week features, which may reflect the different patterns on weekdays and weekends.

3.3.2 Comparison among All Feature Sets

In this part, we compare the effectiveness and efficiency of the selected feature set using RMSE features. The scatter plots of the four sets are shown below, along with their RMSE values.

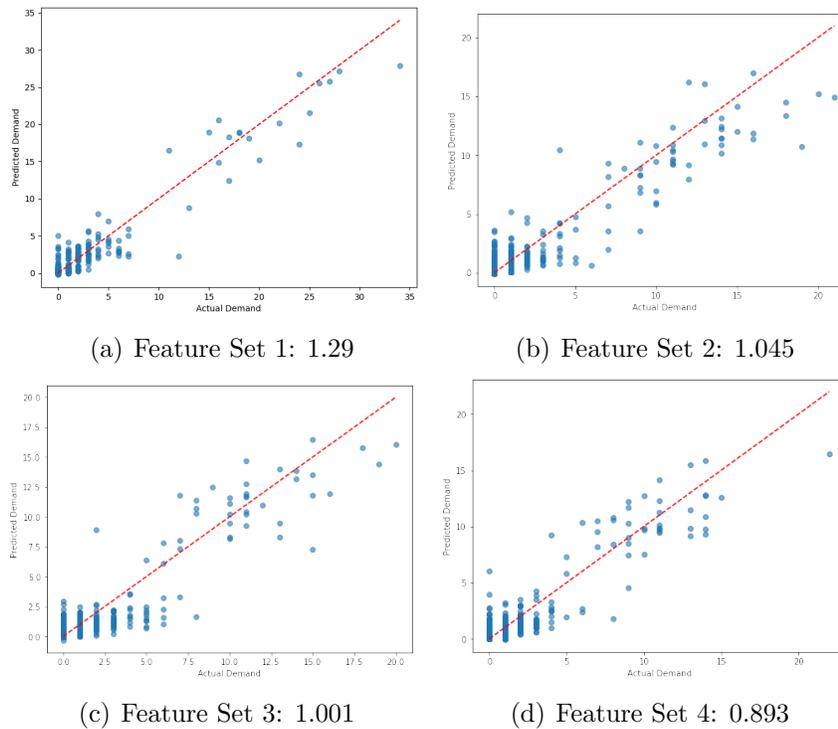


Figure 7: Feature Sets Comparison

From the scatter plots and the RMSE values, we can conclude that the model trained with previous 24-hour data and removing the month-of-year is the most effective. The prediction error is less than 1 user per 30 minutes, which is accurate enough for deployment and can be used as input for optimization model. Also, in the following part, we will use it to make predictions on Caltech charging site data and test its generalization ability.

3.3.3 Generalization Test

In this part, we will test the prediction accuracy of the trained model on a different dataset. Previously, the model was trained on the data collected at the JPL charging station. Now, it will be tested using the data collected from Caltech charging site. The demand pattern is similar but not same. At the same time, we also train a model directly using the Caltech data and compare their performance. The scatter plots of these two models are shown below:

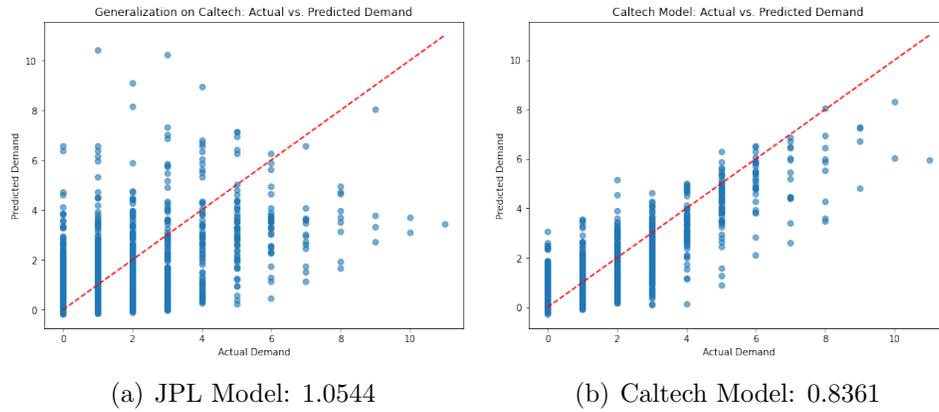


Figure 8: Generalization

We can see that the Caltech model outperforms the JPL model greatly by using the same features. The generalization of the model is poor. When we look into the feature importance ranking of both models, we can see that they value different features. For Caltech, it values hour-of-day since the pattern of the demand in Caltech is more periodical but the JPL sites, the data has higher causality in time series.

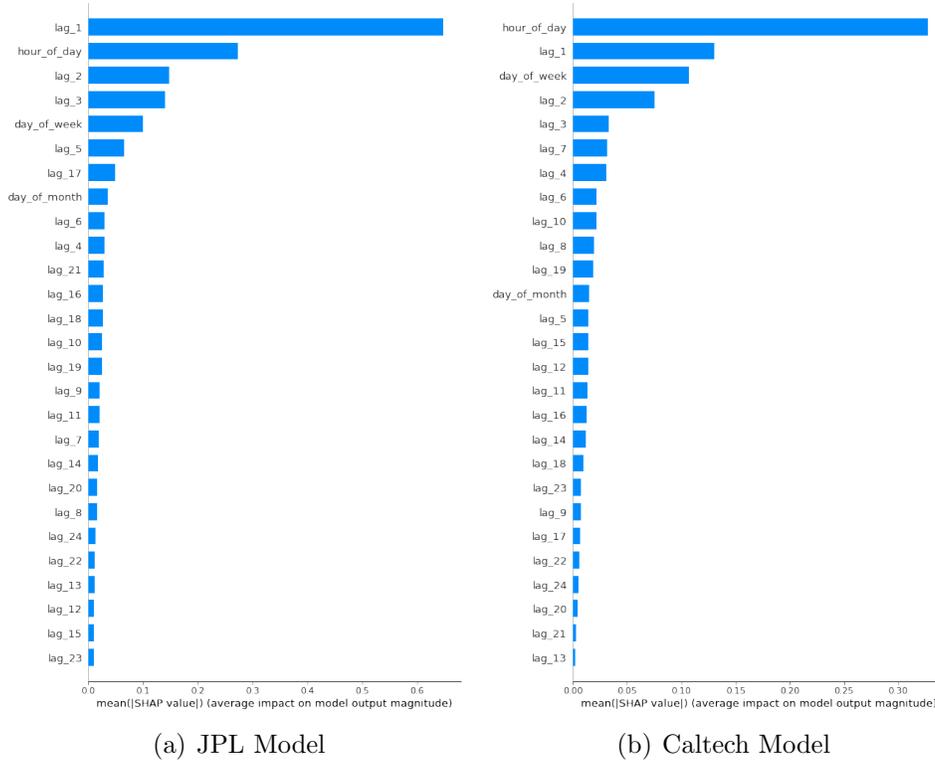


Figure 9: Feature Importance Ranking Comparison

3.4 Summary

The XGBoost model is employed to predict EV charging demand due to its efficiency, scalability, and flexibility in handling large datasets and customized objectives. Using a decision tree ensemble approach, it optimizes predictions by iteratively adding trees and leveraging regularization to prevent overfitting. Feature selection, based on rooted mean squared error (RMSE) and a Shapley Value-based ranking method, identifies the most effective feature

set: 24-hour historical demand without month-of-year features, achieving an RMSE of 0.893. However, a generalization test on Caltech data reveals poor transferability, as feature importance and demand patterns differ from the original JPL dataset. Training a new model on Caltech data achieves superior performance, highlighting the need for tailored models for different datasets. The selected feature set and methodology are effective for deployment and can serve as an input to the optimization module to provide better charging recommendations.

4 EV Charging Station Recommendation System

The EV recommendation system provides suggestions to individual EV users on where to charge and for how long to charge, based on the predicted demand levels, as charging prices are determined by these levels at each station. This suggestion system balances the overall cost of charging with the benefits of charging. Therefore, we propose a mixed-integer programming (MIP) formulation to offer each EV user an optimal recommendation plan that specifies the location, time, and duration for charging.

4.1 Problem Statement

The recommendation system processes requests from EV users, detailing their current location, current time $t_{current}$, current battery level $SOC^{current}$, and the required electric volume $SOC^{required}$. Within an acceptable driving distance range, there exists a set of alternative EV charging stations, denoted by I . The charging start time may allow delays after the request time to take advantage of better prices, referring $t^{start} > t_{current}$. Furthermore, extending the charging duration beyond the required $SOC^{required}$ is advantageous, as longer charging times can reduce the frequency of future charging, thereby lowering the fixed costs associated with charging.

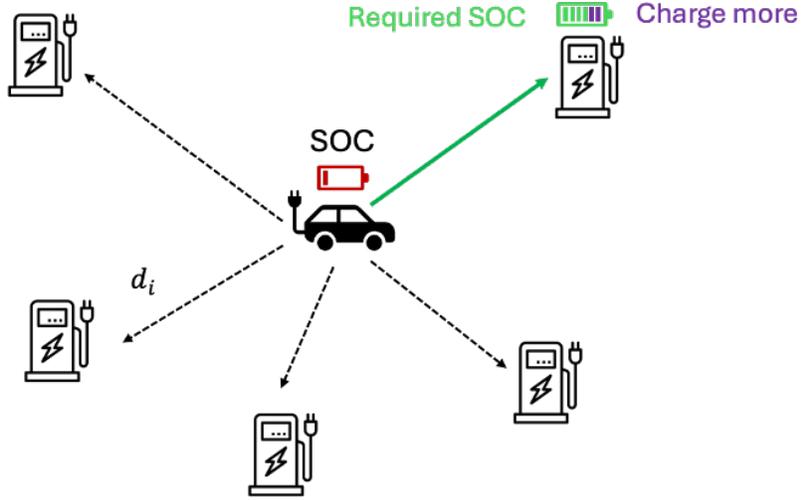


Figure 10: Introduction of the EV Charging Station Recommendation System

We define the charging station decision as a binary variable $x_i, i \in I$, and the charging start and end times as t_{start} and t_{end} , respectively. We divide time into intervals, forming a set T , and define the binary variable w_t to indicate whether the time interval t falls within the charging window, (t_{start}, t_{end}) .

4.1.1 Constraints

The driving distance cannot exceed the limitation imposed by the remaining battery capacity, as expressed by the following constraint:

$$d_i x_i \leq \delta SOC^{current}, \quad \forall i \in I \quad (10)$$

where δ represents the coefficient that converts electric charge into driving distance, and d_i denotes the distance to charging station i .

Equations 11 and 12 indicate that the charging state must remain within the battery limit, \overline{SOC} , while also satisfying the requirement for the required power volume, $SOC^{required}$. In this context, γ serves as the parameter used for converting charging duration into battery volume.

$$\gamma(t_{end} - t_{start}) \leq \overline{SOC} - SOC^{current} \quad (11)$$

$$\gamma(t_{end} - t_{start}) \geq SOC^{require} \quad (12)$$

Equations 13 and 14 indicate whether the time interval w_t falls within the charging time window (t_{start}, t_{end}) . This determination is essential for calculating the price of charging during that specific time window.

$$t_{start} - b_t \leq M(1 - w_t) \quad \forall t \in T \quad (13)$$

$$b_t - (t_{end} - 1) \leq Mw_t \quad \forall t \in T \quad (14)$$

,where b_t represents the start time of time interval t , and M denotes a big number.

4.1.2 Objective Function

The objective function represents the overall cost of charging and encompasses the following components:

- Driving cost from the current location to the charging station.
- Charging price during the designated time window.
- Penalty for late charging due to waiting for a better price.
- Negative benefit from charging more to reduce fixed costs.

The objective function can be formulated as follows:

$$\begin{aligned} \text{obj} = & \alpha^{\text{distance}} \sum_{i \in I} d_i x_i + \sum_{i \in I, t \in T} p_{it} x_i w_t \\ & + \alpha^{\text{late}} (t_{\text{start}} - t_{\text{current}}) - \alpha^{\text{fix}} [\gamma (t_{\text{start}} - t_{\text{current}}) - \text{SOC}^{\text{require}}] \end{aligned} \quad (15)$$

,where α^{distance} , α^{late} , and α^{fix} are monetary parameters that quantify the respective costs associated with each component of the objective function.

4.1.3 Mathematical Formulation

The mixed-integer programming (MIP) formulation for the charging station recommendation system is presented as follows.

$$\begin{aligned} \text{Min } & \alpha^{distance} \sum_{i \in I} d_i x_i + \sum_{i \in I, t \in T} p_{it} x_i w_t + \alpha^{late} (t_{start} - t_{current}) \\ & - \alpha^{fix} [\gamma(t_{start} - t_{current}) - SOC^{require}] \end{aligned}$$

s.t.

$$d_i x_i \leq \delta SOC^{current} \quad \forall i \in I$$

$$\gamma(t_{end} - t_{start}) \leq \overline{SOC} - SOC^{current}$$

$$\gamma(t_{end} - t_{start}) \geq SOC^{require}$$

$$t_{start} - b_t \leq M(1 - w_t) \quad \forall t \in T$$

$$b_t - (t_{end} - 1) \leq M w_t \quad \forall t \in T$$

$$\sum_{i \in I} x_i = 1$$

$$x_i \in \{0, 1\} \quad \forall i \in I$$

$$w_t \in \{0, 1\} \quad \forall t \in T$$

$$t_{start}, t_{end} \geq 0$$

4.2 Case Study and Numerical Result Analysis

We conducted a small case study utilizing 10 alternative EV charging stations over a 24-hour time frame to evaluate the effectiveness of the recommendation system. Figure 11 illustrates the price rates for the ten EV charging stations during this period. The distances to the charging stations are as follows: [1.68, 1.57, 1.69, 1.10, 0.63, 1.06, 1.49, 2.16, 1.70, 2.37] miles.

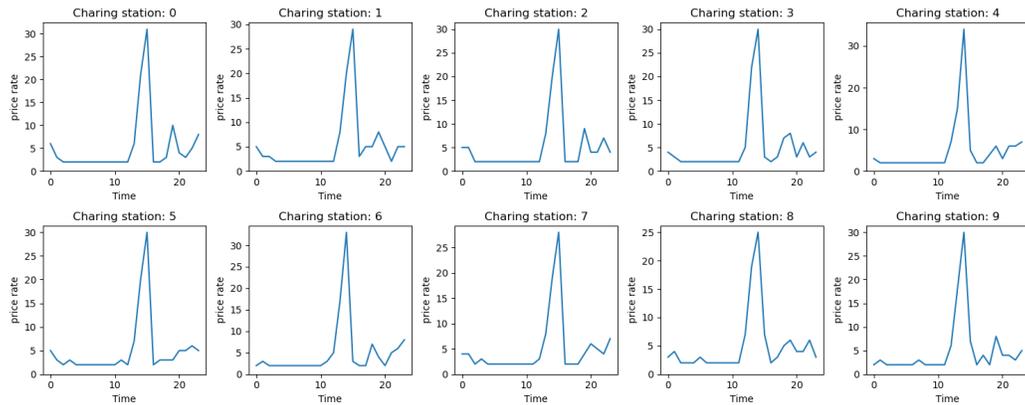


Figure 11: EV Charging Station Price Rates over a 24-hour Period

The recommendation system receives a request at 12:00 AM, with a charging duration requirement of 6 hours. The parameters used in this study are summarized in Table 4.

Table 4: Parameter Value for MIP Formulation

Parameter	Value
γ	10
SOC	85
$SOC^{current}$	2
$SOC^{require}$	60
$t_{current}$	12
δ	3
$\alpha^{distance}$	2
α^{fix}	2
α^{late}	2

Figure 12 illustrates the outcomes of the recommendation system, indicating that charging begins at 4:00 PM at the fourth charging station, with a total charging duration of 8 hours, which exceeds the required charging time by 2 hours.

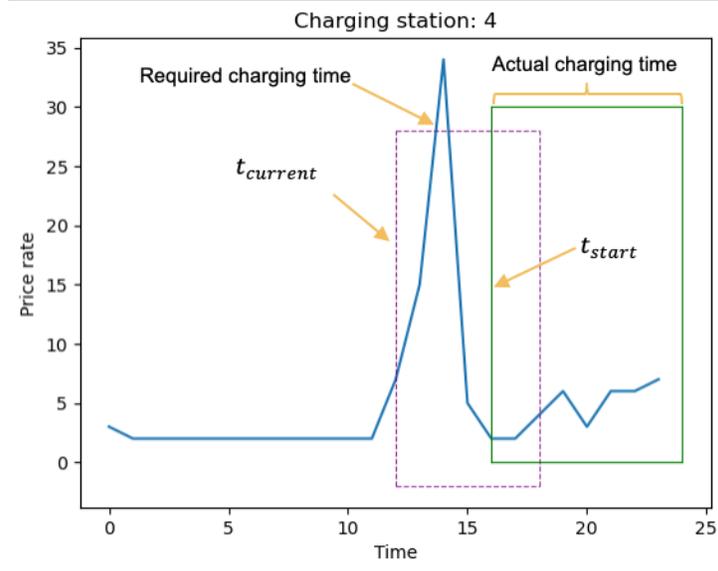


Figure 12: Results of Recommendation System

5 Conclusion

In this project, we developed a comprehensive framework combining demand prediction and optimization to enhance electric vehicle (EV) charging recommendations. The study utilized the Adaptive Charging Network (ACN) dataset, leveraging the exceptional time-series processing capabilities of the XGBoost model for accurate demand forecasting. The final model demonstrated significant improvements over traditional regression approaches.

Despite the model's strong predictive capabilities, our analysis highlighted challenges in generalization across different datasets, as evidenced by varying feature importance rankings and demand patterns between the JPL and Caltech sites. These findings underscore the importance of tailored models for distinct charging environments to ensure accuracy and effectiveness.

The optimization module complemented the predictive model by offering dynamic and personalized charging recommendations, integrating factors such as dynamic pricing, battery states, and user constraints. Our case study demonstrated the system's ability to balance cost-effectiveness, charging efficiency, and user satisfaction, providing a viable solution to manage the growing adoption of EVs.

This project serves as a foundational step toward developing smart and efficient EV charging strategies that promote the advancement and widespread

adoption of EVs. Future research could explore integrating real-time data streams, enhancing model generalizability, and incorporating additional constraints like renewable energy availability. Addressing these aspects will pave the way for more robust and sustainable EV ecosystems, making EVs even more appealing to users.

References

Lee, Zachary J., Tongxin Li, and Steven H. Low (June 2019). “ACN-Data”.
In: *Proceedings of the Tenth ACM International Conference on Future
Energy Systems*. ACM, pp. 139–149. ISBN: 9781450366717. DOI: 10.1145/
3307772.3328313.