XGBoost-Driven Demand Prediction for Optimized Electric Vehicle Charging Recommendations

Enhancing EV Charging Efficiency through Data-Driven Usage Patterns

By: NatDave, Qi, Zhaoming (2024)

Outline

1. Project Overview
2. Data Analysis
3. Prediction Data Flow
4. XGBoost Model
5. XGBoost vs Linear Reg.

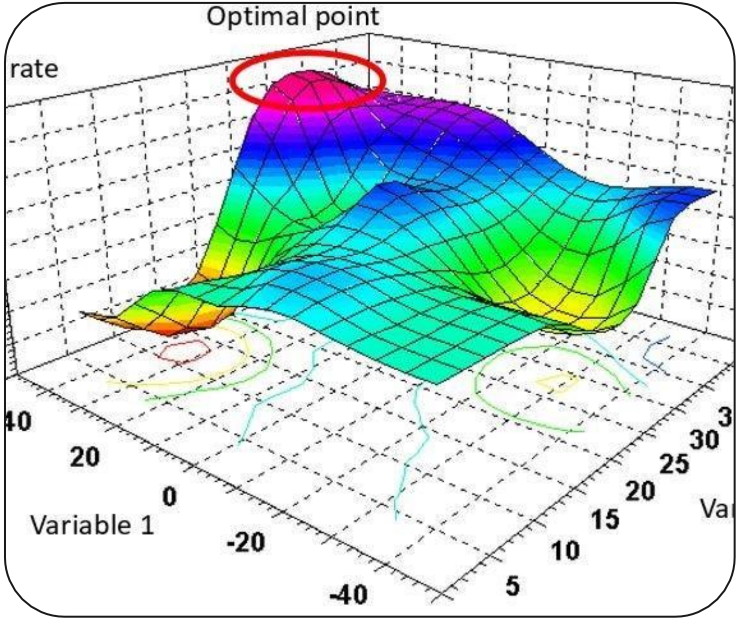
Project Overview

Infrastructure Constraints

Complex EV Charging Decisions

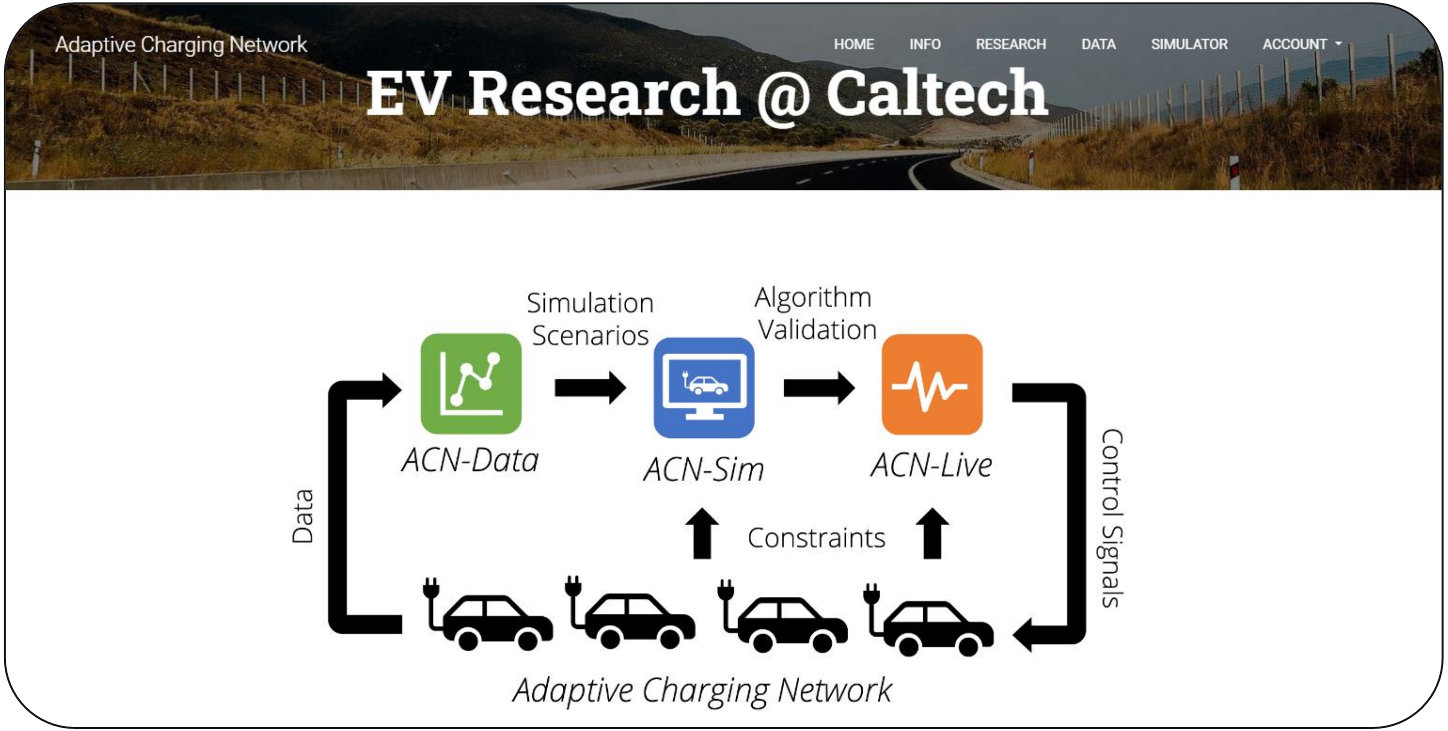
Goals & Expected Contributions

Demand Prediction
(XGBoost)



Charging Recommendations
(Optimization Models)

Dataset



https://ev.caltech.edu

EV Charging Sites

Name	id	Description
Caltech	caltech	A research university located in Pasadena, CA. We currently collect data from 54 EVSEs in one campus garage. The site is open to the public but most usage is from faculty, staff and students.
JPL	jpl	A national research lab located in La Canada, CA. This site currently has 50 EVSEs and is only open to employees. It is indicative of a normal workplace schedule.
Office 1	office001	An office building located in the Silicon Valley area. It currently has 8 EVSEs and is used only by employees.

Some Relevant Features in the Dataset

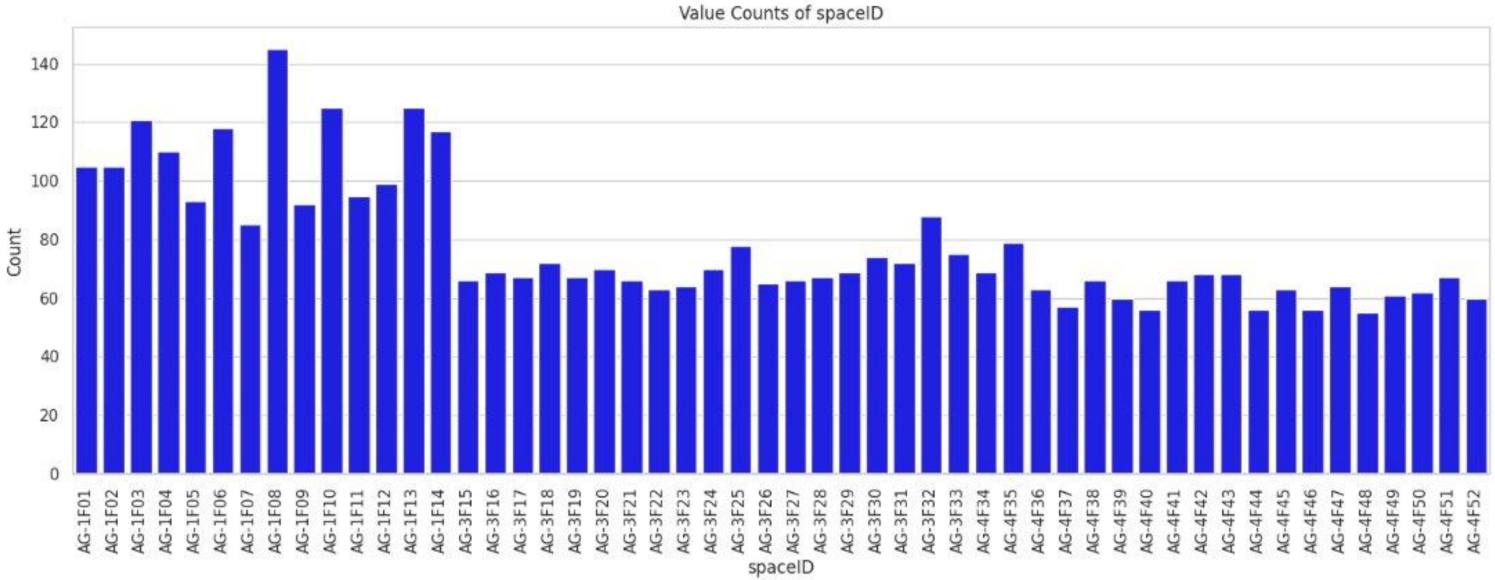
1. **sessionID:** Unique identifier for the session
2. **(dis)connectionTime:** time EV plugged in/out
3. **Time series data:** time zone, timestamp, etc.
4. **kWhDelivered:** Energy delivered during session
5. **User Inputs:** WhPerMile, kWhRequested, etc.

Input Features

Hourly aggregation

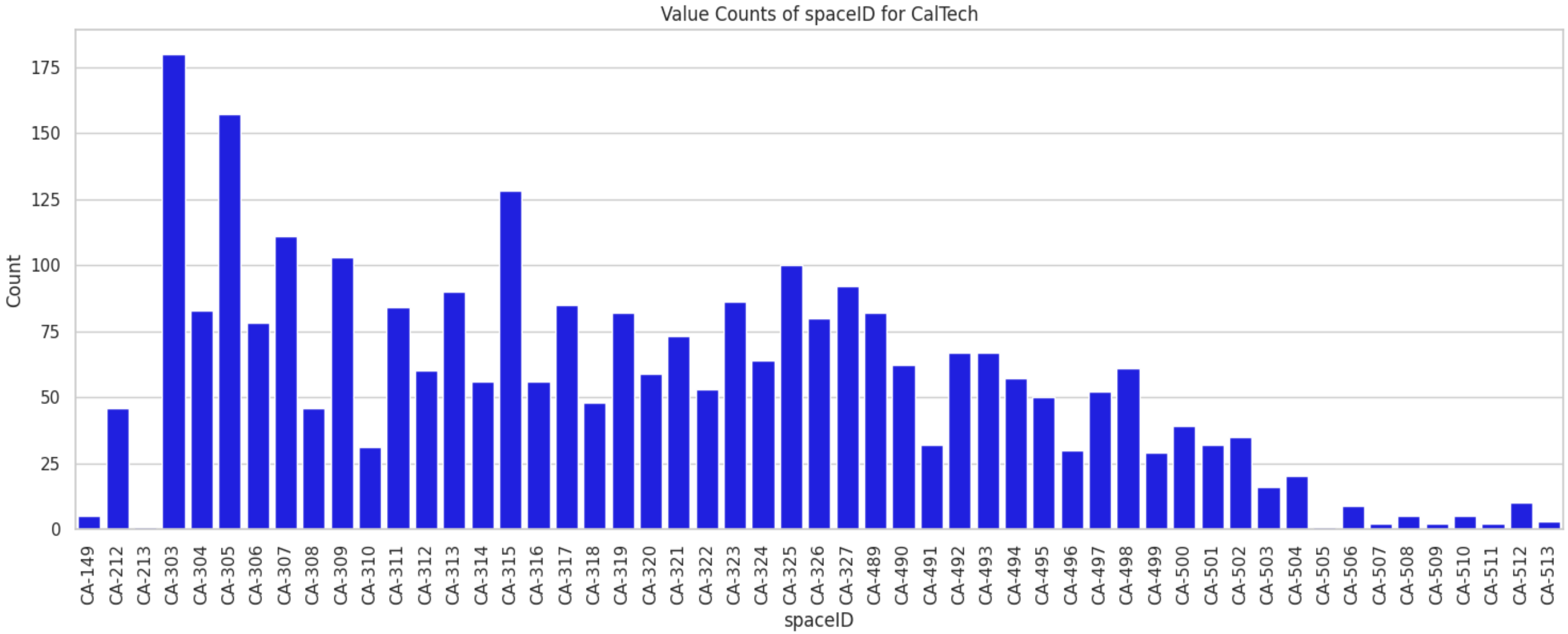
timestamp	lag_1	lag_2	lag_3	lag_4	lag_5	lag_6	lag_7	lag_8	lag_9	lag_10	lag_11	lag_12	hour_of_day	day_of_week	day_of_month	month_of_year	demand
02/01/2019 13:00	0	0	0	0	0	0	0	0	0	0	0	1	13	2	2	1	4
02/01/2019 14:00	4	0	0	0	0	0	0	0	0	0	0	0	14	2	2	1	16
02/01/2019 15:00	16	4	0	0	0	0	0	0	0	0	0	0	15	2	2	1	14
02/01/2019 16:00	14	16	4	0	0	0	0	0	0	0	0	0	16	2	2	1	13
02/01/2019 17:00	13	14	16	4	0	0	0	0	0	0	0	0	17	2	2	1	6
02/01/2019 18:00	6	13	14	16	4	0	0	0	0	0	0	0	18	2	2	1	0
02/01/2019 19:00	0	6	13	14	16	4	0	0	0	0	0	0	19	2	2	1	1
02/01/2019 20:00	1	0	6	13	14	16	4	0	0	0	0	0	20	2	2	1	6
02/01/2019 21:00	6	1	0	6	13	14	16	4	0	0	0	0	21	2	2	1	2
02/01/2019 22:00	2	6	1	0	6	13	14	16	4	0	0	0	22	2	2	1	3
02/01/2019 23:00	3	2	6	1	0	6	13	14	16	4	0	0	23	2	2	1	3
03/01/2019 00:00	3	3	2	6	1	0	6	13	14	16	4	0	0	3	3	1	2
03/01/2019 01:00	2	3	3	2	6	1	0	6	13	14	16	4	1	3	3	1	1
03/01/2019 02:00	1	2	3	3	2	6	1	0	6	13	14	16	2	3	3	1	0
03/01/2019 03:00	0	1	2	3	3	2	6	1	0	6	13	14	3	3	3	1	0
03/01/2019 04:00	0	0	1	2	3	3	2	6	1	0	6	13	4	3	3	1	0
03/01/2019 05:00	0	0	0	1	2	3	3	2	6	1	0	6	5	3	3	1	0
03/01/2019 06:00	0	0	0	0	1	2	3	3	2	6	1	0	6	3	3	1	0
03/01/2019 07:00	0	0	0	0	0	1	2	3	3	2	6	1	7	3	3	1	0
03/01/2019 08:00	0	0	0	0	0	0	1	2	3	3	2	6	8	3	3	1	0
03/01/2019 09:00	0	0	0	0	0	0	0	1	2	3	3	2	9	3	3	1	0
03/01/2019 10:00	0	0	0	0	0	0	0	0	1	2	3	3	10	3	3	1	0
03/01/2019 11:00	0	0	0	0	0	0	0	0	0	1	2	3	11	3	3	1	0
03/01/2019 12:00	0	0	0	0	0	0	0	0	0	0	1	2	12	3	3	1	0

EDA - JPL



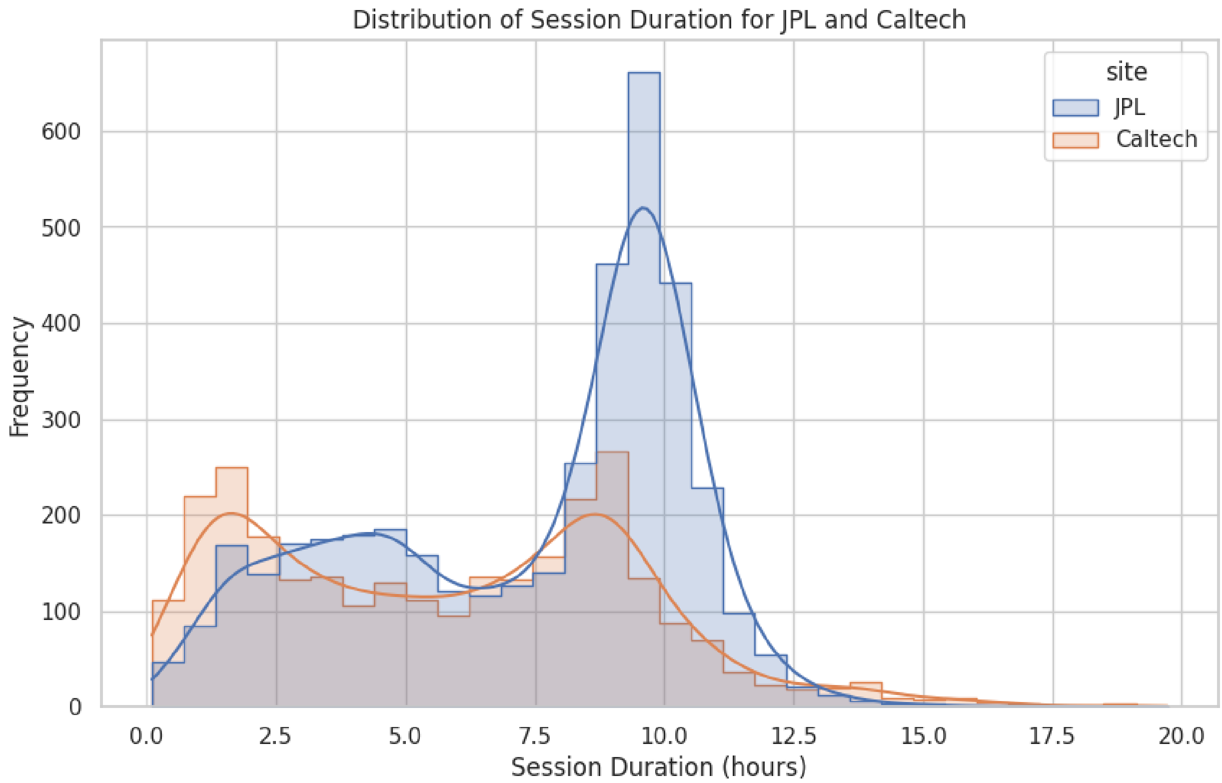
Value Counts of spaceID
Count
spaceID
0
20
40
60
80
100
120
140
AG-1F01
AG-1F02
AG-1F03
AG-1F04
AG-1F05
AG-1F06
AG-1F07
AG-1F08
AG-1F09
AG-1F10
AG-1F11
AG-1F12
AG-1F13
AG-1F14
AG-3F15
AG-3F16
AG-3F17
AG-3F18
AG-3F19
AG-3F20
AG-3F21
AG-3F22
AG-3F23
AG-3F24
AG-3F25
AG-3F26
AG-3F27
AG-3F28
AG-3F29
AG-3F30
AG-3F31
AG-3F32
AG-3F33
AG-4F34
AG-4F35
AG-4F36
AG-4F37
AG-4F38
AG-4F39
AG-4F40
AG-4F41
AG-4F42
AG-4F43
AG-4F44
AG-4F45
AG-4F46
AG-4F47
AG-4F48
AG-4F49
AG-4F50
AG-4F51
AG-4F52

EDA - Caltech

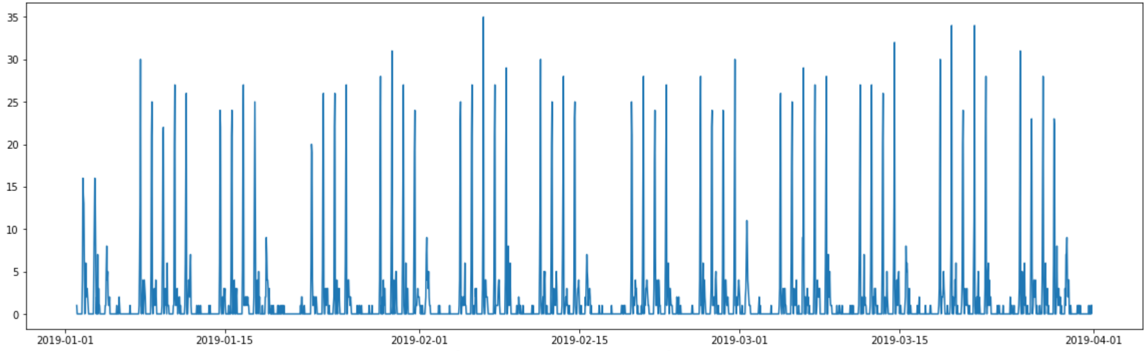


Value Counts of spaceID for CalTech
Count
175
150
125
100
75
50
25
0
CA-149
CA-212
CA-213
CA-303
CA-304
CA-305
CA-306
CA-307
CA-308
CA-309
CA-310
CA-311
CA-312
CA-313
CA-314
CA-315
CA-316
CA-317
CA-318
CA-319
CA-320
CA-321
CA-322
CA-323
CA-324
CA-325
CA-326
CA-327
CA-489
CA-490
CA-491
CA-492
CA-493
CA-494
CA-495
CA-496
CA-497
CA-498
CA-499
CA-500
CA-501
CA-502
CA-503
CA-504
CA-505
CA-506
CA-507
CA-508
CA-509
CA-510
CA-511
CA-512
CA-513
spaceID

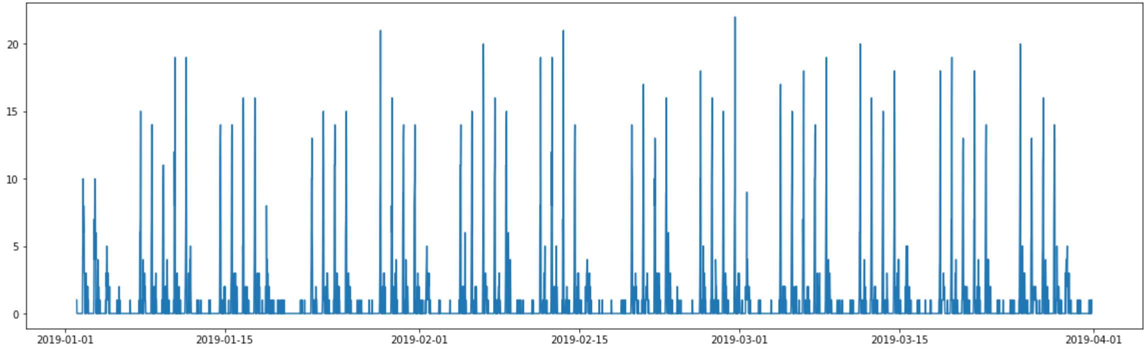
Distribution of Session Durations



Full Hour Against Half Hour Resolution

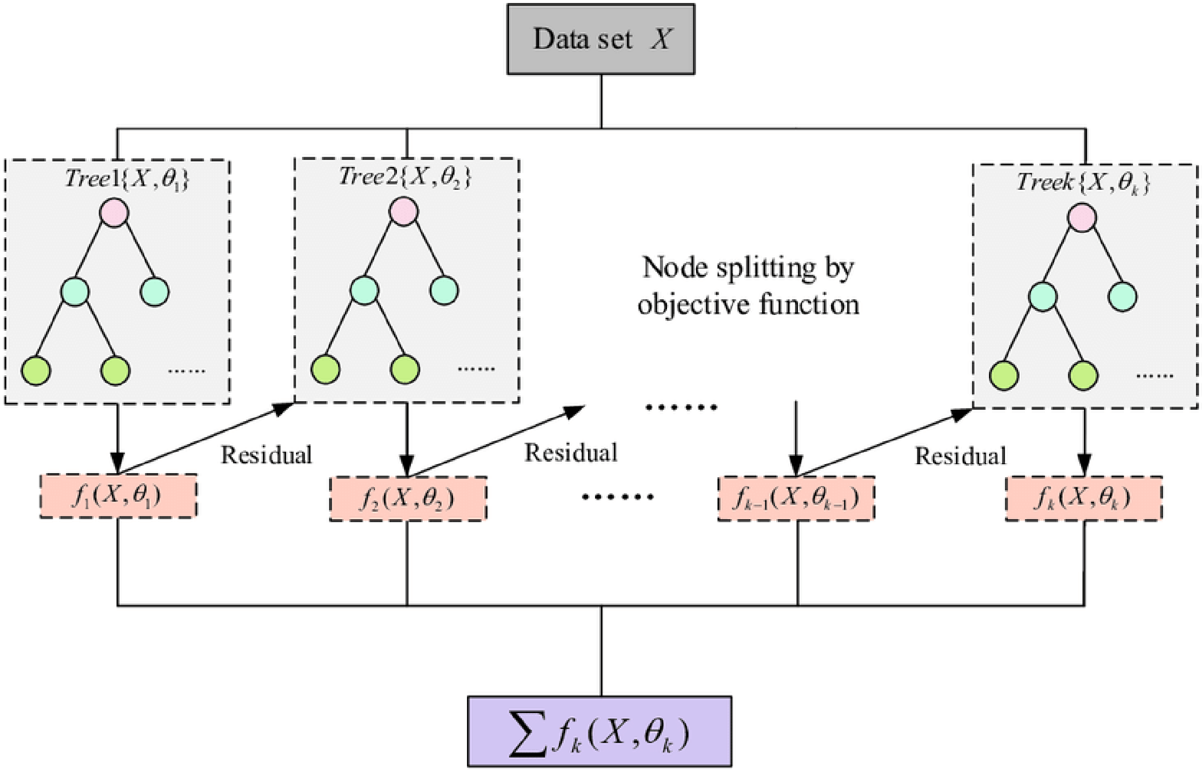


1 hour Resolution



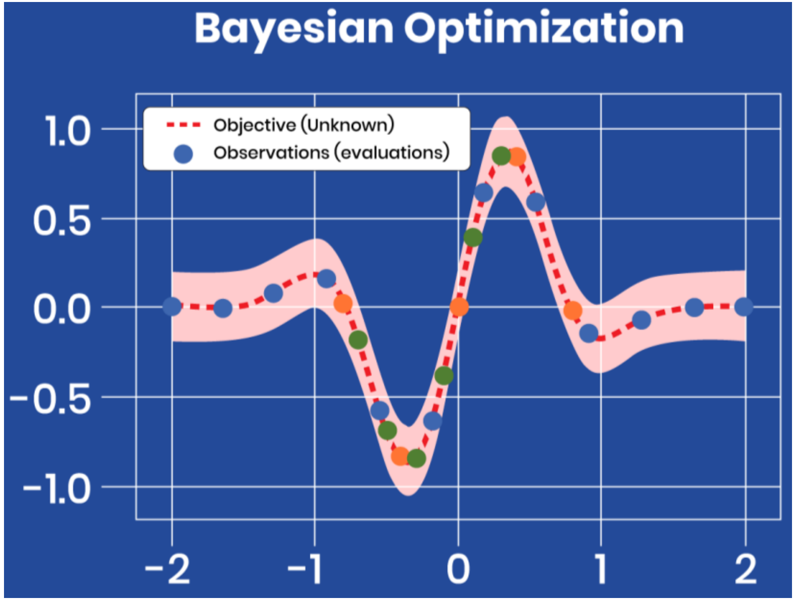
30 minutes Resolution

XGBoost Model Structure

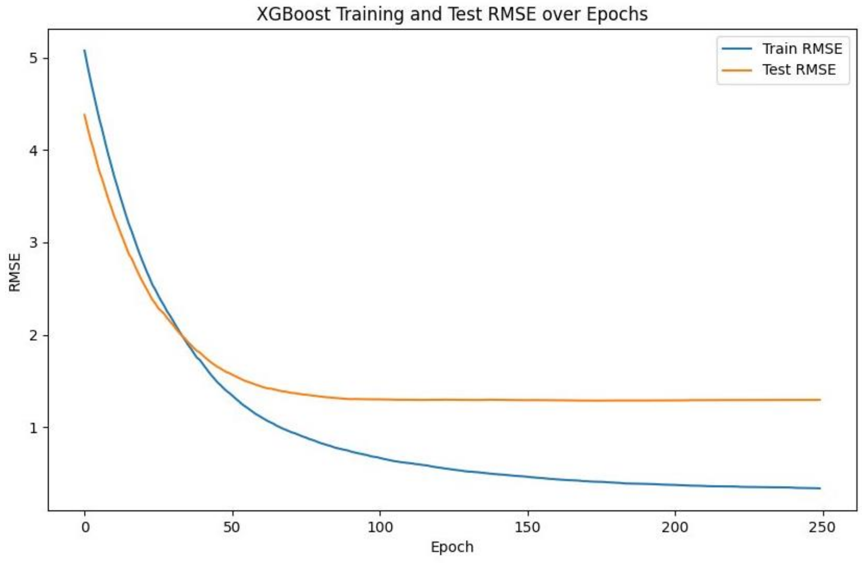


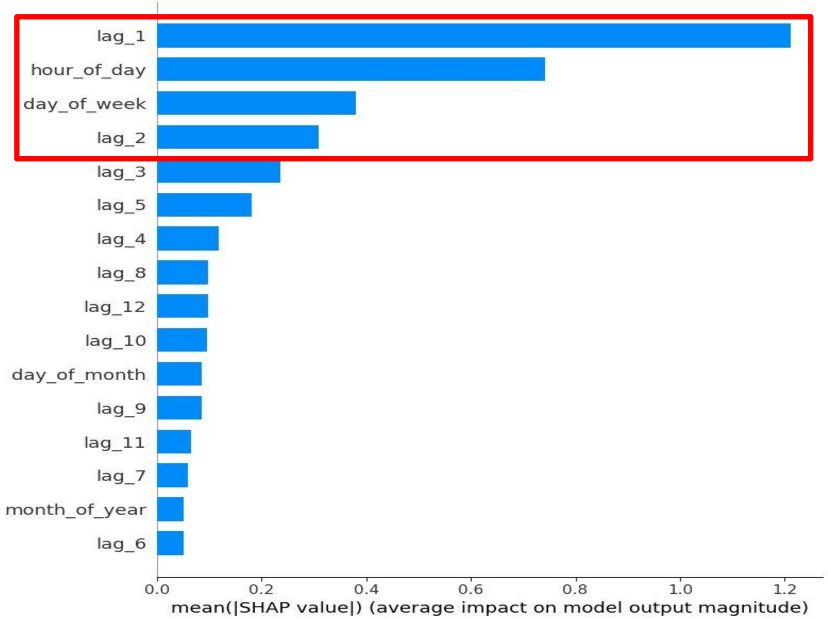
Hyperparameter Optimization (Bayesian)

- ❑ Maximum depth: **8**
- ❑ Child-weight: **1**
- ❑ Sub sample: **.997**
- ❑ Learning rate: **.036**
- ❑ Estimator numbers: **194**

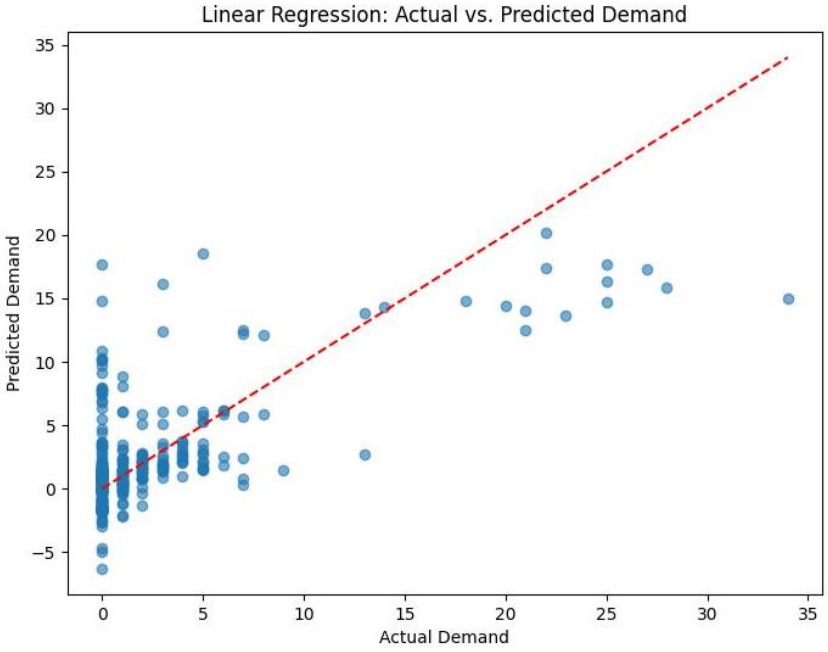


XGBoost Results

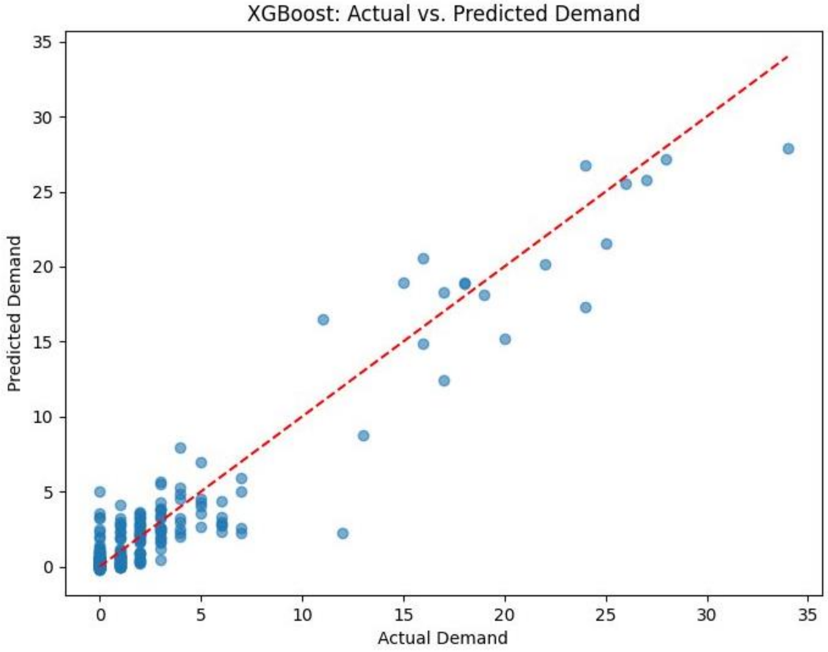




Comparison with Linear Regression

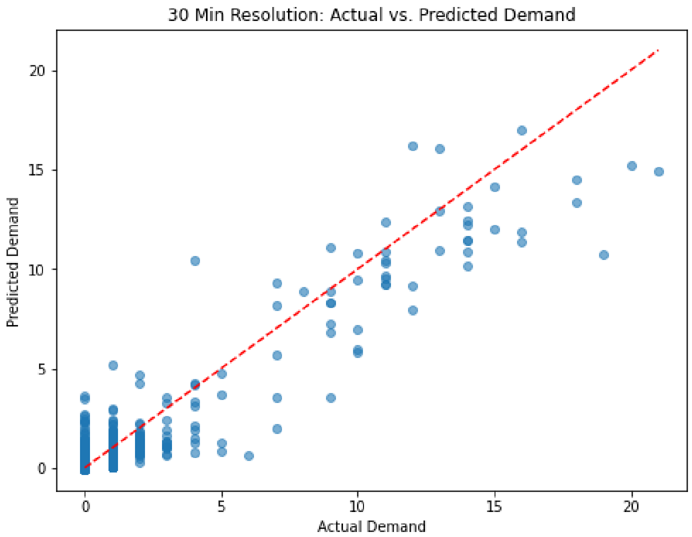


RMSE = 1.82

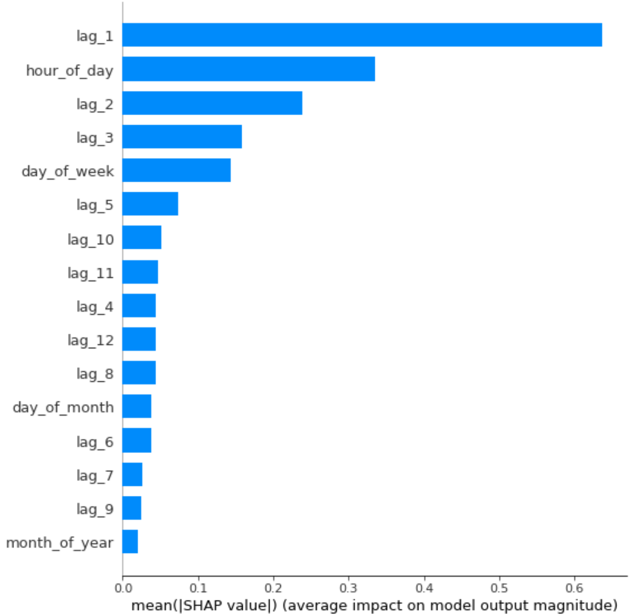


RMSE = 1.29

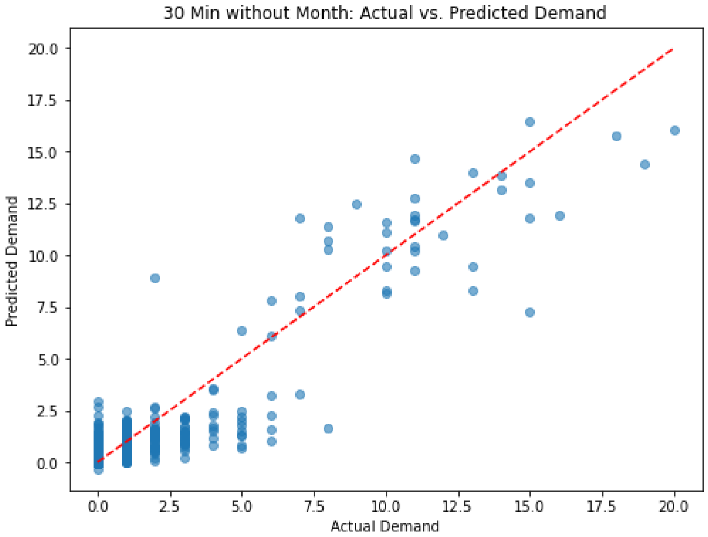
XGBoost 30 min. Resolution



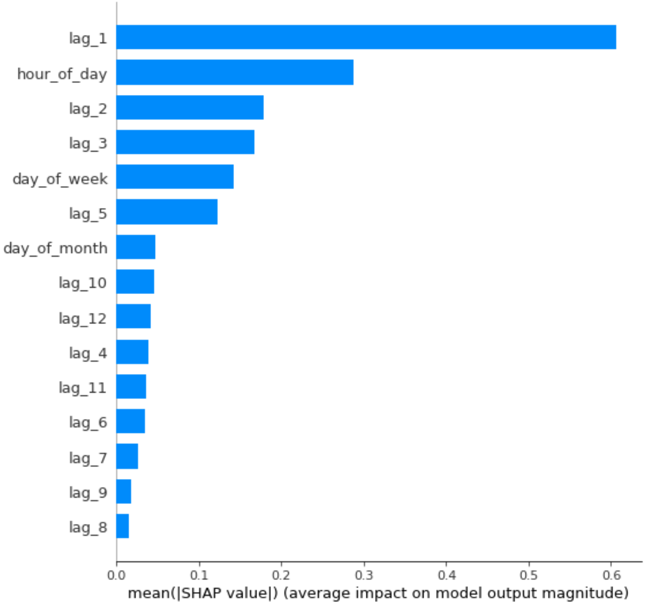
RMSE = 1.045



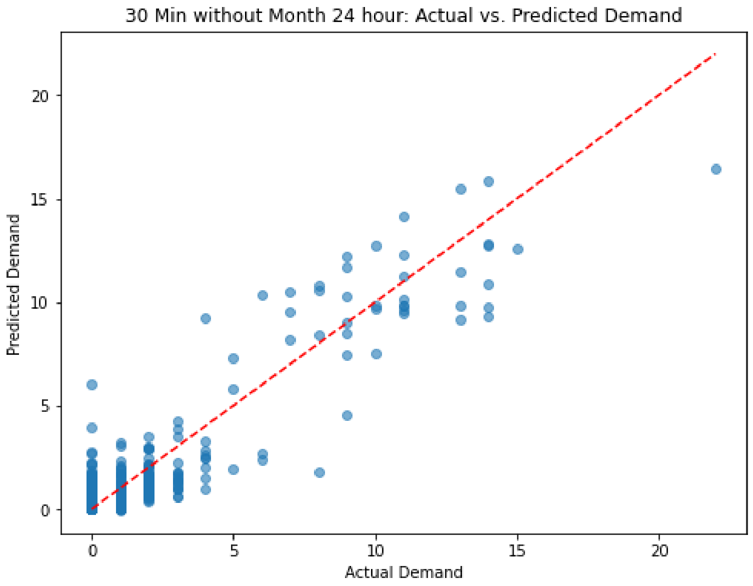
XGBoost 30 min. No Month of Year



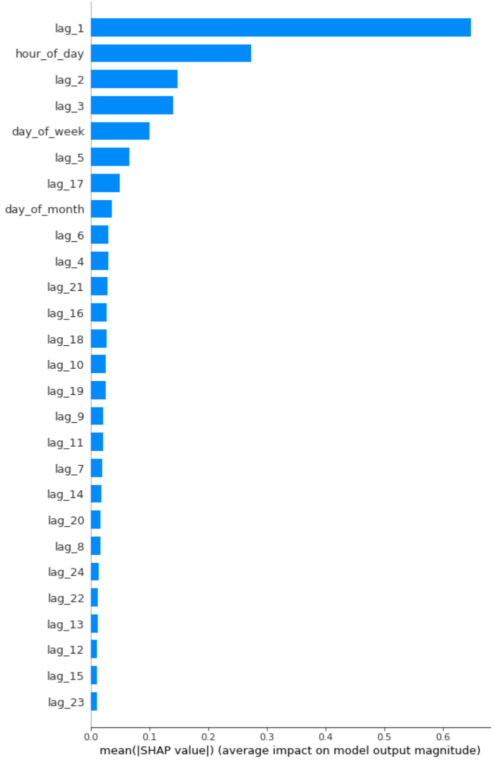
RMSE = 1.001



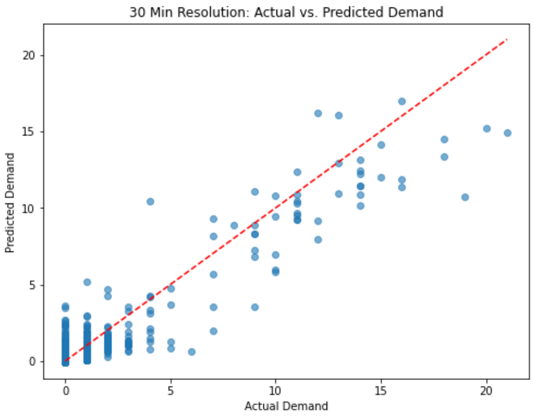
XGBoost 24-hour Input

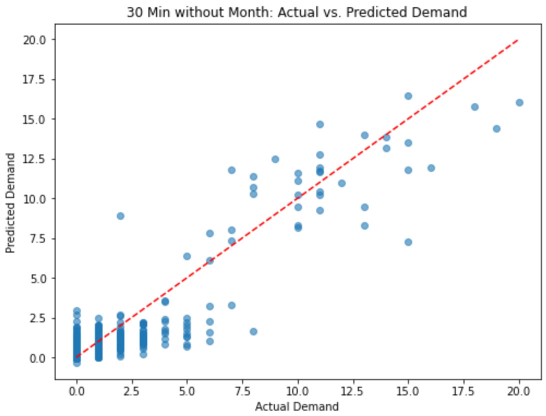


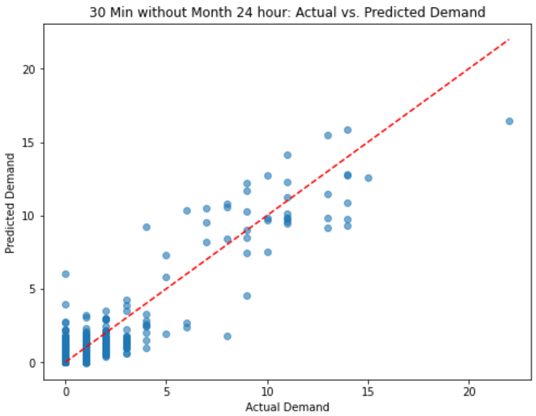
RMSE = 0.893



Scatter Plot Comparison





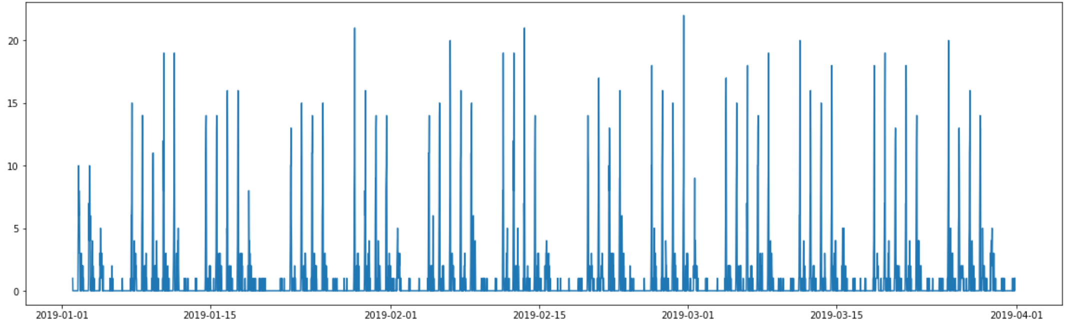


RMSE = 1.045

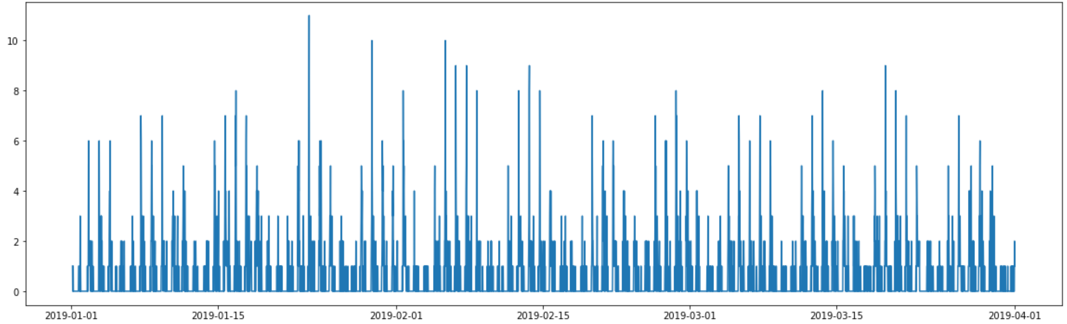
RMSE = 1.001

RMSE = 0.893

Generalization for Caltech Data

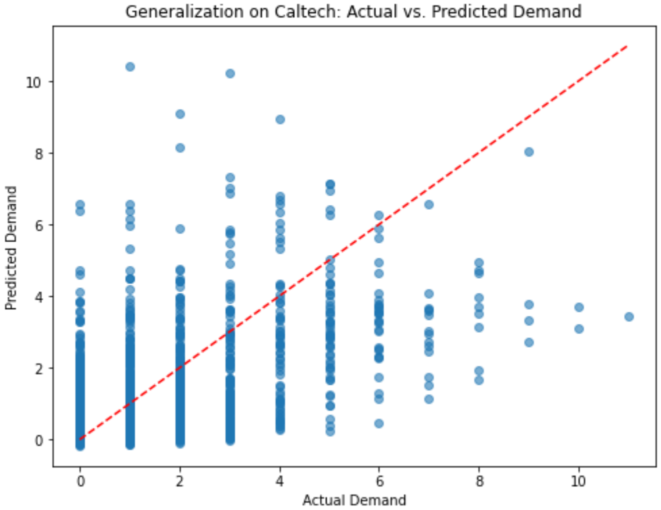


JPL

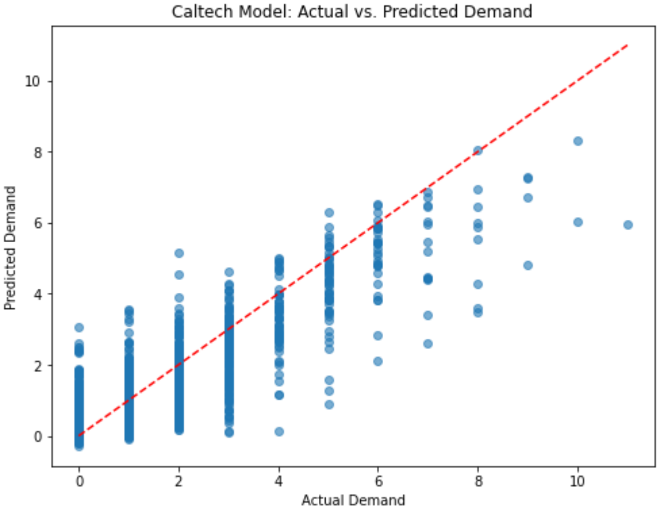


Caltech

Generalization for Caltech Data

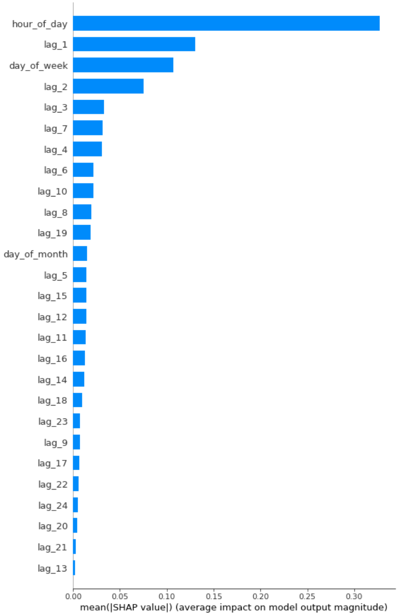


RMSE = 1.0544

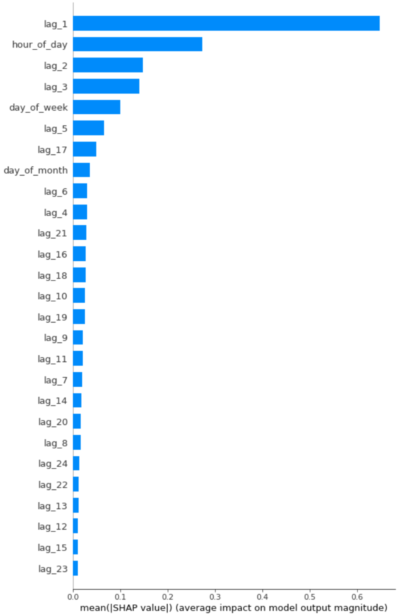


RMSE = 0.8361

Generalization for Caltech Data



Caltech



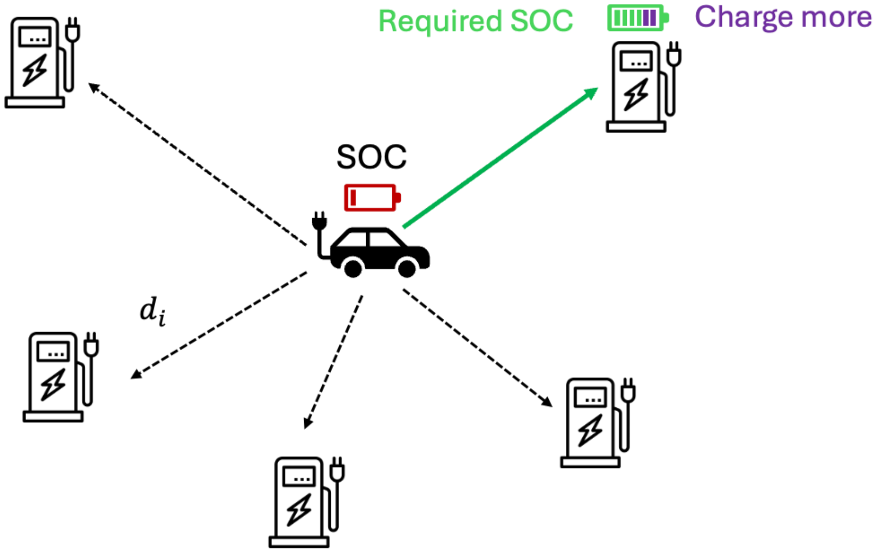
JPL

Charging Recommender System

When and where to charge?

- Ask for charge at $t_{current}$
- Need $SOC^{required}$

- Waiting for good price $t_{start} > t_{current}$
- Charge more than $SOC^{required}$ to get bonus



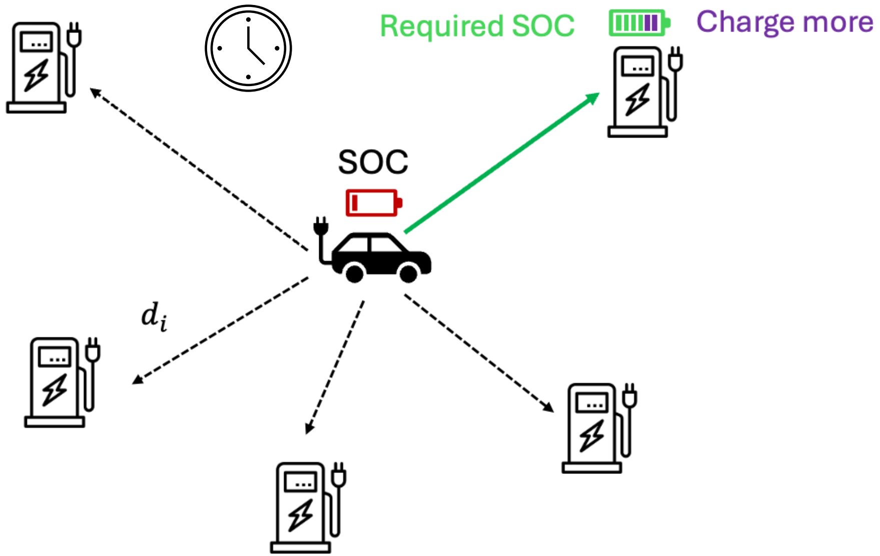
Mixed Integer Programming Formulation

Objective: minimize the overall cost

- ❑ Driving cost
- ❑ Charging pricing
- ❑ Waiting penalty
- ❑ Negative of over charging bonus: reduce charging frequency

Constraints

- ❑ space availability
- ❑ driving distance
- ❑ Charging electricity requirement



Objective: minimize the overall cost

- ❑ Driving cost
- ❑ Charging pricing
- ❑ Waiting penalty
- ❑ Negative of over charging bonus: reduce charging frequency

$$\text{Minimize} \quad \alpha^{\text{distance}} \sum_{i \in I} d_i x_i + \alpha^{\text{price}} \sum_{i \in I, t \in T} p_{it} x_i w_t + \alpha^{\text{late}} (t_{start} - t_{now}) \\ - \alpha^{\text{fixed}} [\gamma (t_{end} - t_{start}) - SOC^{required}]$$

Decision variables

- ❑ Where to charge x_i
- ❑ When to charge t_{start}, t_{end}

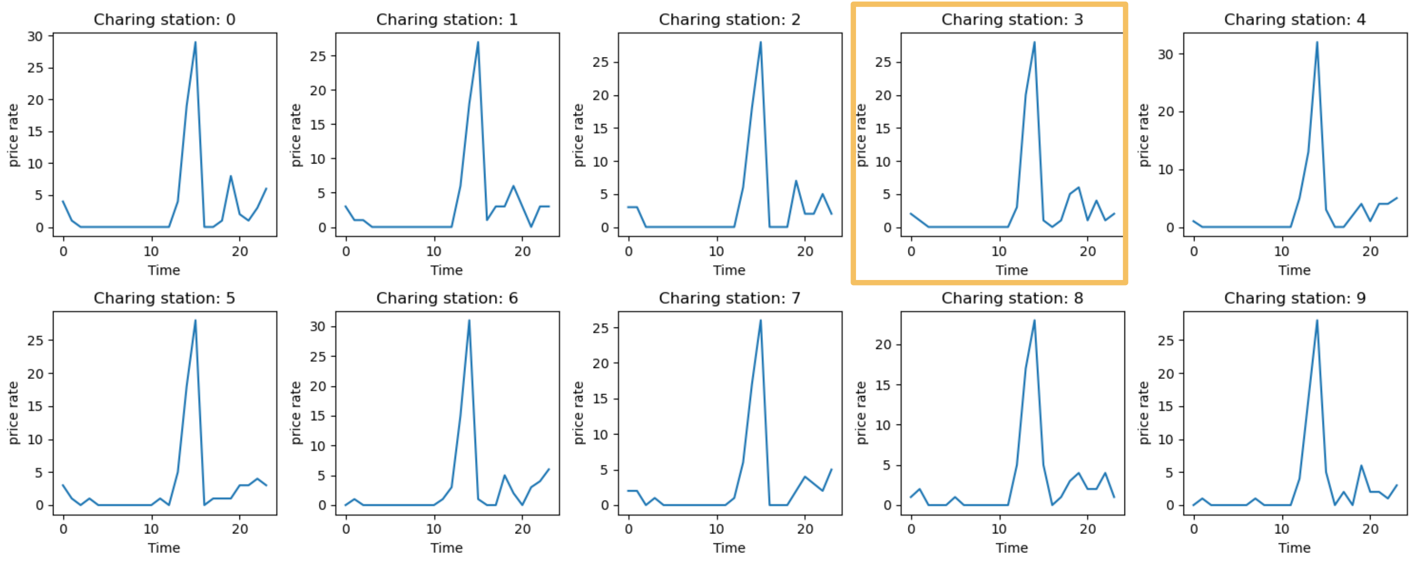
Constraints

Distance requirement	$d_i x_i \leq \delta SOC$	$\forall i \in I$
Electricity requirement	$\gamma(t_{end} - t_{start}) \leq \overline{SOC} - SOC$	
	$\gamma(t_{end} - t_{start}) \geq SOC^{required}$	
Charging time window	$t_{start} - t \leq M(1 - w_t)$	$\forall t \in T$
	$t - (t_{end} - 1) \leq M w_t$	$\forall t \in T$
	$\sum_{i \in I} x_i = 1$	
	$x_i \in \{0, 1\}$	$\forall i \in I$
	$w_t \in \{0, 1\}$	$\forall t \in T$
	$t_{end}, t_{start} \geq 0$	

Case study

Input

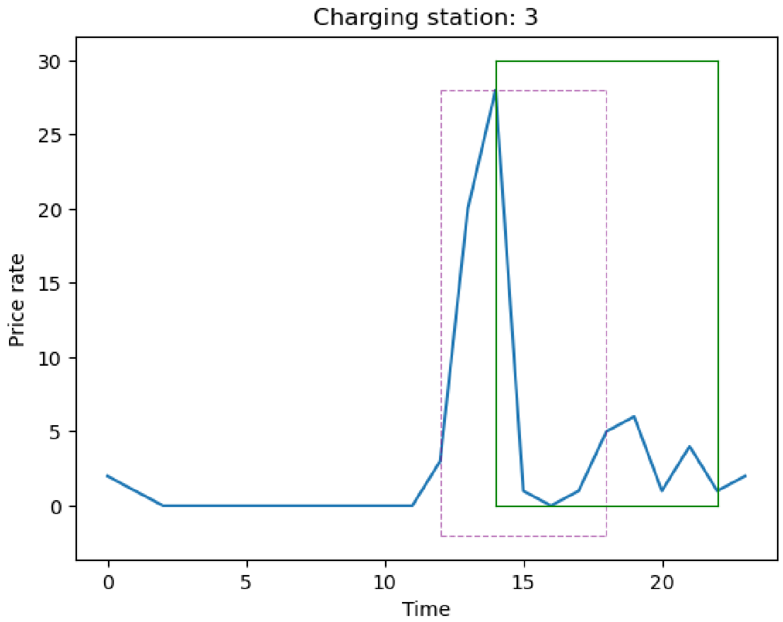
- ❑ 24 hours demand prediction of 10 charging stations
- ❑ Charing request time is 13:00
- ❑ Require 6 hours of charging



Case study

Output

- ❑ Select the 3rd charging station
- ❑ Start at 14:00 and end at 22:00
- ❑ Charging two more hours and wait for two hours



Next Steps?

- ❑ Use **24 hours** (instead of 12) prior data for prediction
- ❑ Predicting **actual** demand or **deviation** from average?
- ❑ **Repeat** test for the two other stations
- ❑ Develop an optimization-based **recommender system**

Questions?