Summer 8-2023

Data-Driven Framework for Predicting and Scheduling Household Charging of EVs

Ahmad Almaghrebi
University of Nebraska - Lincoln

DATA-DRIVEN FRAMEWORK FOR PREDICTING AND SCHEDULING HOUSEHOLD CHARGING OF EVS

by

Ahmad Almaghrebi

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Architectural Engineering

Under the Supervision of Professor Mahmoud Alahmad

Lincoln, Nebraska

August 2023

DATA-DRIVEN FRAMEWORK FOR PREDICTING AND SCHEDULING HOUSEHOLD CHARGING OF EVS

Ahmad Almaghrebi, Ph.D.

University of Nebraska, 2023

Advisor: Mahmoud Alahmad

The increasing prevalence of EV charging poses challenges for power grid stability and quality due to high charging load demands. Without effective energy management strategies for EV charging, the simultaneous power demand from numerous EVs can strain the electric grid, impacting power quality and the wholesale electricity market. To address these challenges, this dissertation presents a comprehensive framework comprising five critical tasks: analyzing EV charging behavior, optimizing charging schedules, developing predictive models, analyzing aggregated impacts, and evaluating implications of predicted user behavior on scheduling. By examining EV charging behavior at household and public charging stations, this study aims to understand patterns and variations in charging sessions. The framework introduces a centralized scheduling approach for household charging stations to reduce peak demand and costs, relying on accurate knowledge of EV charging behavior. Machine learning and linear regression models are utilized to predict session charging parameters, with Random Forest models outperforming other methods, yet uncertainties persist in the predictions. The study also investigates aggregate demand and connectivity of multiple EV users, revealing the potential to predict aggregate trends by incorporating session predictions. Evaluating the day-ahead scheduling framework implemented with predicted charging data using actual data highlights challenges in meeting user demand due to prediction errors. This research provides valuable insights into

EV charging behavior, emphasizing the significance of accurate data for strategic scheduling. While machine learning models show potential in predicting EV charging behavior, limited correlation between session variables and available information at plug-in is observed. Challenges arise from errors in session predictions when scheduling EV charging for a group of users, suggesting the need for a more decentralized approach.

DEDICATION

To my lovely parents Qassem and Salwa and siblings Suhair, Saeed, and Esraa for their continuous love and support.

To my love, Tabarak for her for continuous love, trust and support during the rigorous dissertation work.

ACKNOWLEDGEMENTS

First and foremost, praises and thanks to God, the Almighty, for His showers of blessings throughout my research work to complete it successfully.

I would like to express my heartfelt thanks of gratitude to my research supervisor, Dr. Mahmoud Alahmad Ph.D., PE, for giving me the opportunity to do research and for providing invaluable guidance throughout this research. His dynamism, vision, sincerity, and motivation have deeply inspired me. I would also like to thank him for his friendship and empathy. I am extending my heartfelt thanks to his wife and family for their acceptance and patience during the discussion I had with him on research work.

I would also like to express my special thanks of gratitude to the members of my examination committee, Dr. Fadi Alsaleem, Dr. Iason konstantzos , and Dr.Fabio Vitor, for their inspiring feedback and comments throughout my PhD Journey.

I would like to extend my heartfelt gratitude to Dr. Kevin James. I consider him not only a mentor but also a dear brother. His guidance and encouragement, both academically and personally, have played a pivotal role in my journey and I am deeply grateful for every moment spent with him.

I am profoundly grateful to my loving parents for their unwavering support, encouragement, and sacrifices throughout my academic journey. Your belief in me, guidance, and unconditional love have fueled my determination to overcome challenges and pursue my dreams. I also want to express my heartfelt appreciation to my fiancée, whose trust and support during the rigorous dissertation work have been invaluable. To my dear siblings, thank you for always standing by me with unwavering support, words of encouragement, and a strong bond. Your presence in my life has brought immense joy, laughter, and a sense of belonging. Together, we have created countless cherished memories, and I am forever grateful for the love and support you have shown me.

I am immensely grateful to my mentors, professors, and colleagues who have played a pivotal role in my academic and personal growth, providing me with constant inspiration, invaluable guidance, and intellectual stimulation throughout this extraordinary journey.

I am deeply thankful to my friends in Al-Nabek, Syria, and my friends in Omaha, Nebraska, for their unwavering support and companionship throughout my journey. Your presence and encouragement have been instrumental in helping me achieve my goals, and I am truly grateful for the meaningful friendships we share.

This work has been supported, in part, by the Nebraska Environmental Trust (NET), the Nebraska Community Energy Alliance (NCEA), and the Durham School of Architectural Engineering and Construction (DSAEC) – University of Nebraska-Lincoln.

PREFACE

- The content of Chapter 3 has been published In 2019 IEEE Transportation Electrification Conference and Expo (ITEC). (**Almaghrebi, A**., Shom, S., Al Juheshi, F., James, K., & Alahmad, M.) (pp. 1-6). And in In 2020 IEEE Transportation Electrification Conference & Expo (ITEC) (**Almaghrebi, A.,** Al Juheshi, F., Nekl, J., James, K., & Alahmad, M.) (pp. 1-6).
- Chapter 4 has been published In 2022 IEEE Transportation Electrification Conference & Expo (ITEC) (pp. 540-545).(**Almaghrebi, A**., Vitor, F., James, K., Al Juheshi, F., & Alahmad, M)

Table of Contents

List of Figures

List of Tables

Nomenclature

EVs	Electric Vehicles
IEA	International Energy Agency
RF	Random Forest
ANN	Artificial Neural Network
LOL	transformer Loss of Life
DT	Distribution Transformers
GMMs	Gaussian Mixture Models
SVM	support vector machines
MAE	mean absolute error
MAPE	mean absolute percentage errors
TPR	true positive rate
LSTM	long short-term memory
RNN	Recurrent neural networks
GRU	gated recurrent units
RMSE	root mean square error
NRMSE	normalized root mean square error
HMMs	Hidden Markov Models
KDE	Kernel Density Estimation
DKDE	Dirichlet Kernel Density Estimation
MILP	mixed-integer linear programming
LVDNs	Low Voltage Distribution Networks
SMG	Smart Microgrid
GWO	Grey Wolf Optimization
IBGWO	Improved Binary Grey Wolf Optimization
GA	Genetic Algorithms
PSO	Particle Swarm Optimization
SoC	state of charge
MPC	Model Predictive Control
UCSD	University of California, San Diego
NB	Naive Bayes
LVDNs	Low Voltage Distribution Networks
NPPD	Nebraska Public Power District

1. INTRODUCTION

1.1 Overview

Climate change has emerged as a pressing global issue, prompting numerous efforts to mitigate the effects of global warming [1]. The Paris Agreement of 2015 marked a significant milestone, with countries committing to reducing their emission levels to combat climate change [2]. In line with these objectives, governments worldwide are promoting the adoption of electric vehicles (EVs) as a means to decrease reliance on fossil fuels in transportation [3]. The EV market has witnessed remarkable growth, and this trend is projected to continue in the future [4]. According to the International Energy Agency (IEA), global EV sales reached 6.6 million units in 2021, more than double the 3 million sold in 2020 and over triple the 2.2 million sold in 2019, indicating a steady upward trajectory [5]. Figure 1-1 illustrates the remarkable growth in EV sales over the years.

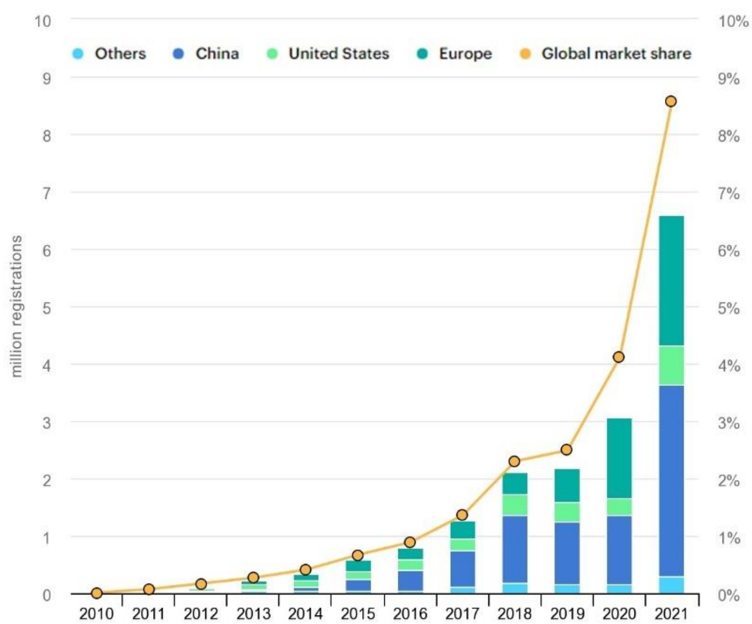

Figure 1-1: Global EV sales and market share 2010 – 2021 [6].

The substantial progress in the adoption of electric vehicles (EVs) can be attributed to the collaborative efforts of policymakers and industry stakeholders in recent years [7]. EVs offer users the advantage of being able to drive using electricity alone or in combination with fuel, making them an attractive option for environmentally conscious individuals. However, the significant increase in EV usage has posed considerable challenges in terms of managing the energy demands and the impact on local energy grid load management. As depicted in Figure 1-2, the rapid growth in EVs has strained energy grids, necessitating effective load management strategies to ensure the stability and reliability of the grid [8]. Addressing these challenges is crucial for maximizing the benefits of EVs while maintaining a sustainable and efficient energy infrastructure.

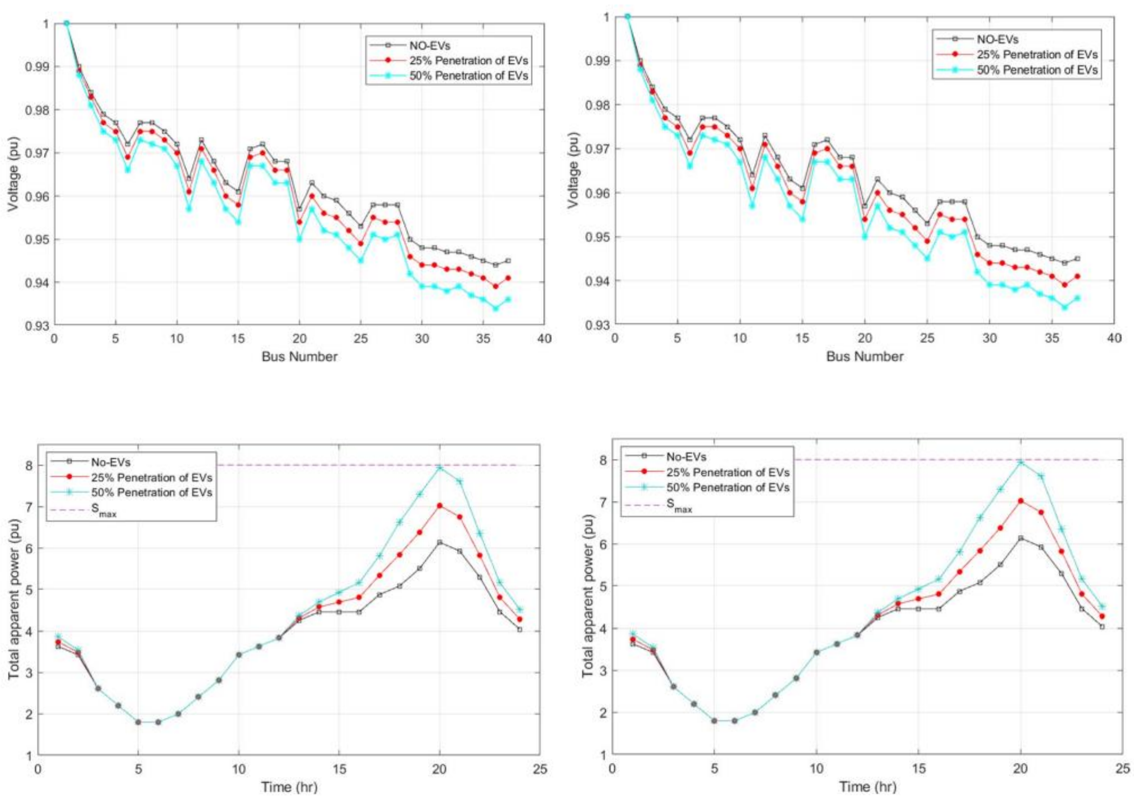

Figure 1-2: The impact of EV penetration on the distribution system [9].

The increasing prevalence of EV charging poses challenges to power grid stability and quality due to high charging load demands. This situation may necessitate costly upgrades to generation capacity and electricity infrastructure. Typically, EV charging loads exhibit two peaks per day, corresponding to morning and evening charging sessions when individuals plug in their vehicles at work and home, respectively. Without effective energy management for EV charging, the simultaneous power demand from numerous EVs can strain the electric grid, leading to lower power quality and potential impacts on the wholesale electricity market. The storage capacity of new EVs, which is approximately 100 kWh, is about four times the daily electricity consumption of an average U.S. household [8], [10]. These factors highlight the urgency of implementing robust energy management strategies to ensure grid stability and mitigate potential negative effects caused by the high demand for EV charging.

Even though an EV battery has a storage capacity of up to 100 kWh, the average daily charging requirement is much lower. Typically, an EV only requires a 10-kWh charge or approximately 1-1.5 hours of charging at a rate of 7 kW using a level 2 charging station, as depicted in Table 1-1. This level of charging is sufficient to meet the average daily driving needs of around 30 miles [11]. It is important to note that most EV owners charge their vehicles overnight, taking advantage of off-peak electricity rates and allowing for a full charge by the morning, ensuring the availability of a fully charged vehicle for daily use.

Table 1-1: Three different power levels of power grid standard of national generating utilities [12].

Power Level Types	**Locations**	**Supply Circuit**	**Power (kW)**	**Fully Charge Duration BEV/ PHEV**
Level 1	Home or Office	120 VAC 1-Phase (20A)	1.4kW at 12A-1.9kW (On-Board)	12-20Hours/7 Hours
Level 2	Private or Public Ports	240 VAC 1-Phase (40-80A)	7.7kW-19.2kW (On-Board)	4-6 Hours/3 Hours
Level 3/DC Fast-Charging	Commercial Stations Like a Filling Station	450VAC/600VDC 3-Phase/DC 200A/400A	6.5 kW-240 kW (Off-Board)	0.5 Hours 10-15 Minutes

The flexibility in EV charging allows for the implementation of optimized charging schedules that can effectively distribute the charging load and mitigate the adverse effects of high demand. There is no immediate need to start charging as soon as an electric vehicle is plugged into a household charging station, as there is typically enough time available for the charging process. By strategically scheduling EV charging, power peaks can be avoided, thereby alleviating the strain on power plants and transformers, without the need for capacity upgrades to the power grid.

Research indicates that a significant number of EV users tend to plug in their vehicles upon returning home from work, resulting in high peak loads that negatively impact power plant and transformer operations. As the adoption of EVs increases, these challenges become more pronounced. A case study depicted in Figure 1-3 illustrates the impact of EV penetration, showing a 43% increase in power consumption and a staggering 98% increase in peak load at 60% EV penetration [13]. These findings underscore the importance of implementing effective charging scheduling strategies to manage the growing demand and ensure the stability and efficiency of the power grid.

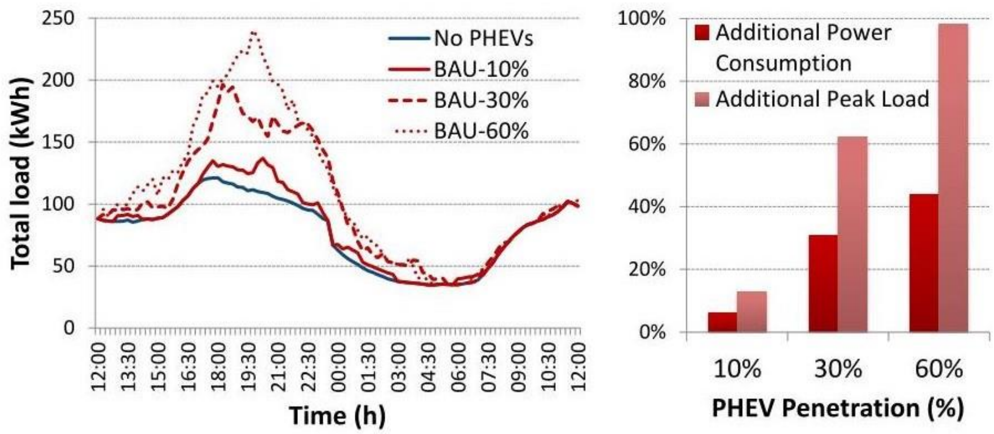

Figure 1-3: Impact of EV penetration on the power [13].

This increase in peak load can be mitigated by shifting some of these charging sessions to off-peak hours. In a controlled charging framework, a utility can directly manage and schedule the charging of all EVs connected to the grid. Such control schemes can drastically reduce the peak demand caused by EV charging, as shown in the case study conducted by Idaho National Laboratory in Figure 1-4 [14].

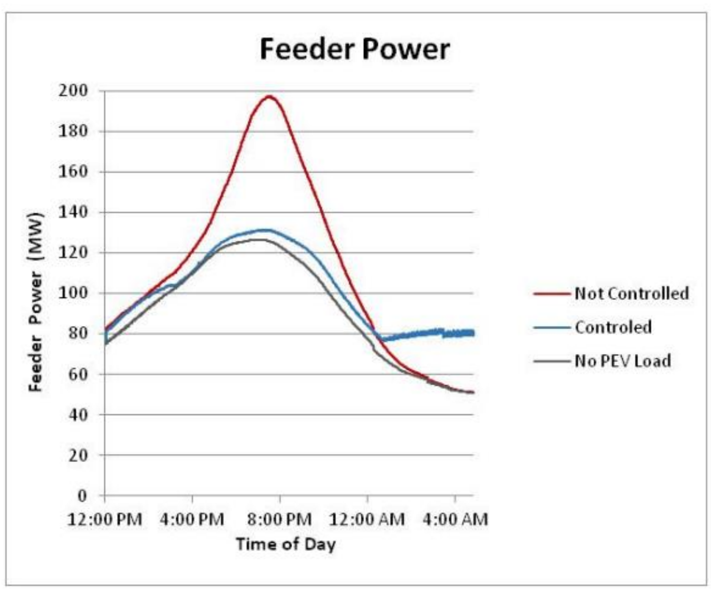

Figure 1-4: Case Study, showing the importance of controlled charging [14].

1.2 Problem Statement

To apply an effective charging schedule for electric vehicles (EVs), utility providers require accurate information about when users will connect and disconnect their vehicles, as well as the amount of energy they will need. However, this information is often uncertain and challenging to obtain directly. Nonetheless, if user behavior follows consistent patterns, it may be possible to predict their charging behavior with a certain degree of accuracy.

Therefore, the ultimate goal of this research is **to leverage real charging data to develop a predictive framework for EV user behavior, and use this framework to inform EV scheduling decisions**.

The following section outlines the individual research questions to be answered in pursuit of this goal.

1.3 Research Questions

The problem statement discussed in the previous section centers around leveraging real charging data to develop a predictive framework for electric vehicle (EV) user behavior and utilizing this framework to inform EV scheduling decisions. Based on this problem statement, the following research questions arise.

1. ***"What patterns exist in the behavior of EV users?"*** By analyzing real charging data, an in-depth understanding of the usage patterns exhibited by EV users can be obtained. This includes examining the timing and duration of charging sessions,

connection durations, and energy consumption. Identifying consistent patterns will establish a foundation for predicting future charging behaviors.

2. ***"How can real charging data be leveraged to optimize EV charging schedules and maximize efficiency?"*** Through a comprehensive analysis of EV users' charging behavior, valuable insights can be derived regarding their preferences, needs, and the factors influencing their charging decisions. This information will guide the development of an effective scheduling routine that optimally aligns charging sessions with user behavior, grid constraints, and other relevant factors. Ultimately, this will enhance the efficiency and reliability of EV charging infrastructure.

3. ***"How can EV charging characteristics be accurately predicted?"*** Building upon the analysis of usage patterns and behavior, one of the key objectives is to develop accurate predictive models for forecasting the charging demand of electric vehicles (EVs). This entails leveraging relevant features, including historical charging data and user-specific characteristics, to train machine learning models. By utilizing these models, it becomes possible to predict various aspects of EV charging sessions, such as connection duration, charging duration, energy consumed, and time until the next session.

4. ***"What are the aggregate impacts of EV charging behavior in an area, and how well can these be predicted?"*** Analyzing the total energy demand and identifying simultaneous connections from aggregated charging data of EV users within a specific geographic area provides valuable insights into the charging behavior. This

analysis allows for a better understanding of the energy requirements and the occurrence of multiple EVs charging simultaneously.

5. ***"If EV charging is scheduled based on predicted behavior, how well can it satisfy actual user demand?"*** Evaluating the performance of the scheduling optimization algorithm is essential to ensure that the scheduled charging sessions align with the actual timeline and meet users' scheduling preferences. This assessment involves examining the algorithm's ability to minimize conflicts, maximize user satisfaction, and optimize the utilization of charging infrastructure. By analyzing these factors, the scheduling process for EV charging can be refined and improved.

1.4 Dissertation Framework

To address the research questions and achieve the objectives of this study, a comprehensive research framework is proposed in Figure 1-5.

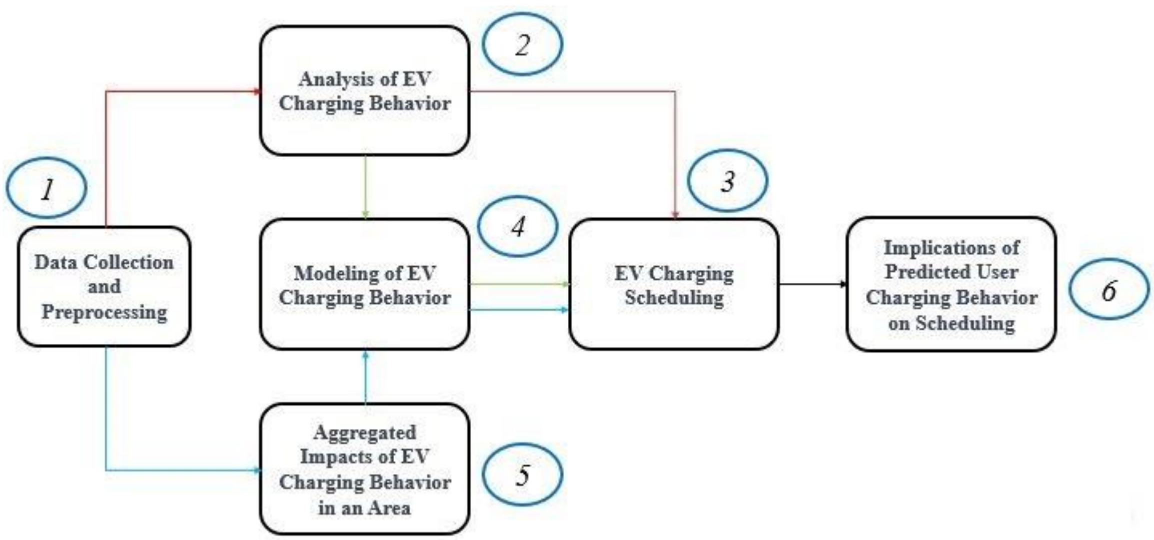

Figure 1-5: Dissertation framework.

The framework encompasses six critical tasks that are essential for leveraging real charging data to develop predictive models of EV user behavior and inform EV scheduling decisions. These tasks involve data collection and preprocessing, analysis of EV charging behavior, scheduling of EV charging, modeling of EV charging behavior, Aggregated impacts of EV charging behavior in an area, and implications of predicted user charging behavior on scheduling.

1. Data Collection and Preprocessing

The data collection process involved gathering charging data from a range of Level 2 charging ports, specifically single-phase 40A and 240V charging stations. These charging stations were selected from both household and public charging infrastructure, ensuring a diverse dataset that encompasses different charging scenarios and user behaviors. This comprehensive dataset provides valuable insights into EV charging patterns and behavior. The collected information includes important features such as start time, end time, connection duration, charging duration, idle duration, and energy consumption, enabling a detailed analysis of EV charging behavior.

For the household charging stations, the dataset consists of 275,827 charging sessions obtained from 485 charging stations located in Omaha, NE. The data spans from April 2018 to December 2022, capturing a significant period of charging activity in residential areas.

Similarly, for the public charging stations, the dataset comprises 84,577 charging sessions obtained from 160 charging stations located in Nebraska. The data covers a longer timeframe, from January 2014 to December 2022, providing a comprehensive view of EV charging behavior in public locations.

2. EV Charging Behavior Analysis

Analysis of EV charging behavior has predominantly focused on public charging stations, yet understanding home charging behavior is equally important for utility companies. The analysis entails examining various aspects, including the start and end times of charging sessions, connection duration, charging duration, idle duration, and energy consumption.

3. EV Charging Scheduling

Building upon the analysis of EV charging behavior, a deeper exploration of the relationship between the identified charging patterns and the scheduling of EV charging sessions is crucial. This investigation aims to uncover the potential impacts and challenges associated with scheduling EV charging, especially considering the anticipated increase in the number of EVs connected to the electric grid. This section explores a framework for the centralized scheduling of household charging stations, using real unscheduled charging data as a baseline for user behavior. A single-day scheduling model is presented, and potential improvements in electricity cost and peak load are examined at multiple levels of EV penetration.

4. EV Charging Behavior Modeling

The EV Charging Behavior Modeling section focuses on the process of developing predictive models using machine learning techniques. The modeling process involves several key steps, including data preprocessing, feature engineering, model selection, and model training and evaluation. In this study, three machine learning algorithms, namely Random Forest (RF), XGBoost, and Artificial Neural Network (ANN), along with Linear Regression, were utilized to predict four important outputs: connection duration, charging duration, energy consumption, and the time until the next charge. These models leverage the collected charging data to capture complex relationships between various input features and the desired output variables.

5. Aggregated Impacts of EV Charging Behavior in an Area

The analysis of electric vehicle (EV) user behavior within a specific geographical area is a crucial aspect to explore after gaining insights from EV charging demand prediction at the session level. This section focuses on studying charging demand and the number of simultaneously connected vehicles in a specific region or community. By analyzing EV user behavior, valuable insights can be gained regarding peak demand periods and the utilization of charging infrastructure. The analysis includes examining the distribution of charging events across various timeframes, such as day, week, or year, to identify peak demand periods and potential grid stress. Session-level models are employed to predict charging demand and the number of simultaneously connected vehicles in the studied region.

6. Implication of Predicted User Charging Behavior on Scheduling

To achieve a practical and effective charging schedule, the integration of predicted data is crucial. This section focuses on incorporating session predictions, such as connection duration and charging duration, into the deterministic scheduling process. The schedule's effectiveness will be assessed based not only on its ability to minimize peak demand and lower electricity costs, but to meet the actual charging demand of EV users.

1.5 Advancements in the State of Research - Contributions of the Study

This study significantly advances the research on EV charging scheduling optimization by presenting a framework for centralized scheduling of household charging stations. The framework aims to minimize peak demand, reduce electricity costs, and meet the charging demands of EV users. Through an analysis of the trade-offs between peak load reduction and cost optimization, the study sheds light on the complexities of scheduling charging sessions. Moreover, the research leverages predictive modeling techniques, including random forest, XGBoost, artificial neural networks (ANN), and linear regression, to accurately predict crucial charging parameters such as connection duration, charging duration, energy consumption, and time until the next charge. By examining charging behavior at both household and public charging stations, valuable insights into charging patterns and preferences are obtained. By incorporating predicted user behavior into the scheduling process, the study enables more efficient resource allocation and enhances the overall efficiency of the charging infrastructure. These advancements have significant implications for grid management, cost efficiency, and user satisfaction, thereby facilitating the wider adoption of electric vehicles and fostering sustainable transportation systems. The contributions of this study can be summarized as follow:

- This study addresses a significant gap in the existing literature by incorporating household charging data, which has been relatively understudied compared to public charging data. By including this unique dataset, the research provides valuable insights into EV charging behaviors in residential settings. This fills an important gap in the literature and contributes to a more comprehensive understanding of the impact of residential charging on the local electricity grid.

- This study introduces a novel set of input features for the predictive modeling of EV charging sessions. By extracting additional features from the historical behavior of each user, including the cumulative mean, minimum, maximum values of the outputs, and number of charging sessions, the study aims to enhance the categorization and prediction of charging behavior. These features provide valuable insights into user charging patterns and contribute to improving the accuracy of the predictive models.
- This study predicts EV charging behavior without relying on unique user IDs as a specific variable. By taking a broader approach and focusing on the statistics of arbitrary users rather than individualized relationships, the study provides insights into the general patterns and dynamics of EV charging behavior.
- This study focuses on modeling the aggregated impacts of EV charging behavior within a specific area. By considering the collective effects of multiple users charging simultaneously, the research aims to understand the broader implications on the local electricity grid and associated challenges. The developed predictive models provide insights into the overall charging patterns,

peak demand periods, and utilization of charging infrastructure in the given area.

- This study proposes a framework for the centralized scheduling of household charging stations, which aims to address the challenges of peak demand management and cost optimization while meeting the charging demands of individual users. The novelty of this framework lies in its ability to consider the specific requirements and preferences of each user while optimizing the overall charging schedule. By integrating predictive models, optimization algorithms, and user preferences, the framework provides a comprehensive and efficient solution for coordinating household charging activities to achieve a balanced reduction of both peak load and cost.
- This study incorporates predicted user behavior into the scheduling of charging sessions, leveraging the developed predictive models for charging behavior. This approach contributes to more realistic and proactive scheduling strategies that align with the expected behavior of EV users. Additionally, the study quantifies the influence of prediction accuracy on charge scheduling, exploring the impact of prediction quality on the effectiveness of the scheduling outcomes.

1.6 Dissertation Organization

The dissertation is organized as follows:

Chapter 1- Introduction

Chapter 2- provides a comprehensive literature review that explores the existing body of knowledge related to EV charging behavior and scheduling. The literature review covers topics such as EV charging behavior analysis, predictive modeling techniques, scheduling algorithms, and optimization approaches.

Chapter 3- In this chapter, an analysis of EV user charging behavior at both household and public charging stations is conducted using real data obtained from accessible Level 2 charging ports in the state of Nebraska.

Chapter 4- In this chapter, a framework for the centralized scheduling of household charging stations is proposed using real unscheduled charging data as a baseline for user behavior with the goal to minimize peak load and electricity costs while satisfying user demand.

Chapter 5- In this chapter, a novel data-driven framework is proposed to predict the EV charging behavior after a charging session starts. This approach is validated using a dataset consisting of five years of charging events collected from household charging stations in Omaha. Nebraska.

Chapter 6- In this chapter, the simultaneous charging of multiple EV users is examined. The analysis focuses on total energy demand and simultaneous connections, using both actual and predicted session behavior.

Chapter 7- This chapter focuses on evaluating the effectiveness of the scheduling optimization algorithm in aligning the charging sessions of electric vehicles with the actual timeline and satisfying users' scheduling preferences.

Chapter 8- This chapter serves as the conclusion of the dissertation, where the findings are summarized, and recommendations for potential future research directions are presented.

2. LITERATURE REVIEW

The literature review on EV charging behavior provides a comprehensive analysis of the existing research aimed at understanding, predicting, and optimizing the charging behavior of electric vehicles. This body of work explores various aspects related to EV charging, predictive modeling, and scheduling optimization. The studies reviewed employ diverse methodologies, ranging from data-driven approaches to machine learning models, to analyze and predict key factors such as charging duration, connection duration, arrival and departure times, and charging demand. These models utilize a wide range of input features, including historical charging data, temporal factors, user preferences, and charging infrastructure characteristics. The findings highlight the complex interplay between EV users, charging infrastructure, and the power grid, underscoring the need for accurate predictions and intelligent scheduling algorithms to mitigate peak demand, manage grid stability, and optimize charging efficiency.

2.1 Impact of Electric Vehicles on Electric Grid

The increasing adoption of electric vehicles (EVs) brings both opportunities and challenges to the electric grid. While EVs offer environmental benefits, their charging behavior can have a significant impact on grid stability and reliability. The rise in electricity demand for charging, especially in low-voltage systems, can lead to conflicts and affect the lifespan of transformers [15].

Studies have shown that the introduction of EVs can result in a considerable increase in transformer Loss of Life (LoL). For example, in [16], a 10X increase in LoL was observed when EVs were introduced, and the annual LoL in urban areas could increase from 0.002 to 0.014. Interestingly, the charging scenario, whether slow or fast charging, has a contrasting effect on power equipment strain. Slow charging, typically done at home during peak afternoon hours, puts more strain on power equipment compared to fast-charging during off-peak hours.

Furthermore, EV usage can accelerate the aging of Distribution Transformers (DT), as analyzed in [17] for an apartment complex with EV chargers. Realistic EV charging demand profiles were generated, and it was found that DT aging could be expedited by up to 40% with an EV penetration ratio of up to 30%. The study also highlighted the potential benefits of integrating PV sources to enhance DT reliability.

It is important to consider realistic charging profiles when assessing the impact of EV load on the grid. In [18], the effects of real charging profiles, particularly peak demand, were examined to gain a better understanding of where and how charging occurs. This dynamic analysis provides insights into potential challenges and opportunities associated with EV charging.

The concerns regarding EV charging impacts were reiterated in [19], where it was shown that frequent charging throughout the day can significantly affect the performance of distribution transformers. Even with a low number of EVs in the transportation sector, the addition of more public fast chargers can lead to transformer overloading.

Uncoordinated EV charging patterns can significantly worsen the impact on the electric grid, necessitating costly upgrades and investments. This uncontrolled behavior strains the infrastructure, potentially leading to increased electricity prices and grid instability. To successfully integrate EVs into the grid, accurate understanding and prediction of charging behavior are essential. This knowledge helps identify challenges, develop efficient strategies, and mitigate the impact on the grid. Intelligent charging scheduling and optimization techniques enable distribution of the charging load, aligning it with low-demand periods and maximizing renewable energy utilization.

2.2 Understanding EV Charging Behavior

Research on electric vehicle (EV) user behavior has made significant contributions to understanding the adoption and usage of electric vehicles. These studies have explored various factors influencing EV adoption, including charging infrastructure availability, socio-economic factors, policy incentives, and socio-cultural and environmental factors. Surveys and interviews have been utilized to capture EV user perceptions and experiences, while quantitative methods such as data analysis and modeling have examined charging patterns and identified influencing factors. The choice of methodology depends on research objectives, available data, and the desired level of detail. Mixed-methods studies that integrate quantitative and qualitative approaches provide a comprehensive understanding of EV user behavior by capturing both statistical patterns and contextual nuances. Overall, the existing literature on EV user behavior has provided valuable insights into the complex dynamics of EV adoption and usage, informing strategies and policies for the widespread adoption of electric vehicles.

In [20], researchers analyzed charging data from a large-scale EV fleet and identified distinct clusters of charging behavior based on user profiles, charging locations, and durations. This analysis provided insights into peak charging periods and opportunities for load balancing and optimizing charging infrastructure placement. Similarly, authors in [21] conducted a comprehensive analysis of EV user behavior in an aggregated area, examining charging data from residential, workplace, and public charging stations. The study highlighted the importance of understanding charging patterns for effective load management and grid stability during peak demand periods.

A different perspective was taken in [22], where researchers explored the impact of pricing structures on EV user behavior. They investigated the response to time-of-use pricing and observed a shift towards off-peak charging. The study demonstrated the potential of pricing mechanisms to incentivize desired charging behaviors and support grid management objectives.

User satisfaction and preferences in EV charging were the focus of [21]. Through surveys and user feedback analysis, factors influencing satisfaction were identified, such as charging speed, convenience, and availability of charging stations. This study emphasized the importance of considering user-centric design in developing charging infrastructure and services.

Future charging demand estimation was addressed in [23], where predictive models were developed to forecast EV charging demand. By considering factors such as EV adoption rate, user behavior, and charging infrastructure growth, the models provided insights for grid planning and infrastructure expansion.

Real and simulated datasets have been employed to analyze EV charging behavior in various studies. For example, [24]–[26] analyzed real data from public charging stations to identify correlations between EV driver behavior and charging patterns. The start time of charging sessions was found to be related to energy consumption and session duration. In [27], a dataset of 400,000 charging sessions in Netherlands was collected, and a Monte Carlo model was applied for trip data simulation.

Authors in [28] employed probability density functions based on Gaussian Mixture Models (GMMs) to determine EV charging metrics. The analysis of users' charging behaviors to study the hourly electricity demand profile was the focus of [29]. An algorithm was developed to predict changes in EV charging demand over time, considering the time and location of charging sessions.

In [30], [31] a load profile for charging EVs was generated using information from travel surveys, taking into account similarities between EV and conventional vehicle travel behavior. Charging behavior concerning start time and location was found to be influential in several studies, such as [24], [25], [32], and [33]. EV charging behavior on weekdays and weekends was analyzed in [33] using charging station and travel data from six European countries to predict the electricity capacity required for charging EVs. Challenges in the electric network resulting from EV charging were predicted in [34] using data from charging points, tracking the charging and travel behavior of real EV users over more than two years.

The majority of research on EV charging behavior has predominantly focused on public charging stations, overlooking the significance of home charging behavior for utility companies. However, with the expected surge in EV adoption, understanding residential charging patterns becomes crucial due to the potential strain it can impose on residential grids. Although existing studies have provided valuable insights into EV charging behavior, there remains a need for a comprehensive analysis and modeling approach to enhance our understanding of EV user charging behavior.

To bridge this gap, this study aims to undertake an in-depth analysis of EV charging behavior by leveraging a diverse dataset collected from both public and household charging stations. By incorporating data from these two sources, the study can capture a wider range of charging scenarios and user behaviors, providing a more comprehensive understanding of EV charging patterns.

2.3 Predictive Modeling of EV Charging Behavior

Predictive modeling of EV charging behavior has been a topic of interest in recent years, with researchers exploring various approaches and methodologies to develop accurate models [26], [35]–[41]. Supervised machine learning techniques, such as decision trees, random forests, support vector machines (SVM), and artificial neural networks (ANN), have been commonly employed in these studies.

One study by [42] used machine learning algorithms, including XGBoost, to predict the departure time of EVs based on a dataset of over 100,000 charging sessions. The model achieved a mean absolute error (MAE) of 82 minutes for departure time prediction. Another study [43] focused on predicting the arrival and departure times of EVs using a dataset from UCSD, with mean absolute percentage errors (MAPE) of 2.85% for arrival time prediction and 3.7% for departure time prediction

Ensemble models were utilized in [37] to predict whether EVs in a household would be charged the next day and the hours of charging. Random Forests (RF), Naive Bayes (NB), and Artificial Neural Networks (ANNs) were combined to achieve a high true positive rate (TPR) and accuracy for these predictions. In a different approach, [44] employed mean estimation and logistic regression to predict user behavior in terms of start time, session duration, and energy consumption. The predictions aimed to stabilize the power grid, although the performance was not evaluated.

Regression models have also been utilized for EV charging behavior prediction. In [26], XGBoost outperformed linear regression, random forest (RF), and support vector machine (SVM) models in predicting energy requirements using public charging stations data. The study achieved a high R^2 score and mean absolute error (MAE) on the test set.

Some studies have gone beyond analyzing user behavior and attempted to predict various charging outcomes. For example, [45] proposed a model to represent the common behavior of EV drivers using real EV data, exploring the statistical characteristics of charging duration, vehicle connection duration, and EV demand profile. The study highlighted the impact of behavioral parameters on congestion status at charging stations.

Anticipating EV charging demand and addressing resulting challenges have also been explored. [46] designed an urban fast-charging demand forecasting model based on a data-driven approach and human decision-making behavior. The model effectively predicted the spatiotemporal distribution characteristics of urban fast-charging demands. [47] developed a data-driven method to predict the popularity of charging infrastructure using Geographic Information Systems data.

Finally, [35] compared different regression methods for determining the idle time of vehicles using data from Netherlands, finding that XGBoost produced the most accurate predictions.

In addition to supervised machine learning techniques, deep learning models have indeed gained significant attention in the predictive modeling of EV charging behavior, leveraging their ability to capture complex patterns and dependencies in data. Recurrent neural networks (RNN), long short-term memory (LSTM), and gated recurrent units (GRU) are commonly used deep learning architectures in this domain, enabling the modeling of temporal dynamics and non-linear relationships.

One area where deep learning models have been successful is electric load forecasting, which is crucial for energy management and operation. In the case of EV charging, short-term load forecasting is particularly important for smart scheduling, considering the arrival and departure times of EV drivers. In [48], deep learning techniques, including LSTM-based models, were utilized for forecasting charging load at shorter intervals, showing superior performance compared to traditional artificial neural network (ANN) models in terms of root mean square error (RMSE) and mean absolute error (MAE).

Hourly charging load prediction at public charging stations was investigated in [49], where multiple RNN-based models, including LSTM and GRU, were employed. The GRU model with one hidden layer achieved the best performance, with a normalized root mean square error (NRMSE) of 2.89%. The study demonstrated the effectiveness of deep learning models in capturing the complex patterns in charging load data.

For super short-term intervals, such as minute-level data, deep learning models have also been applied. In [36], LSTM outperformed conventional ANNs for load forecasting, reducing the forecasting error by more than 30%. The LSTM model achieved the best performance with a mean absolute error (MAE) of 0.29 kW and a root mean square error (RMSE) of 0.44 kW, highlighting the ability of deep learning to handle high-resolution time series data.

Deep generative methods, such as in [50], have been employed to classify charging profiles of EVs based on the distribution of charging arrival and departure times. These methods have demonstrated better performance compared to benchmark models like Hidden Markov Models (HMMs), indicating the effectiveness of deep generative models in capturing charging behavior patterns.

Clustering techniques, such as k-means clustering, hierarchical clustering, expectation-maximization (EM) algorithm, Gaussian Mixture Models (GMMs), and DBSCAN, have proven valuable in modeling EV charging behavior. These techniques allow for the identification of distinct groups or clusters within charging data, providing insights into charging patterns and preferences without the need for explicit labels or target variables.

K-means clustering has been used to categorize user charging behavior into different groups based on features like arrival and departure times, as well as the correlation between stay duration and energy consumption. This approach was employed in [51], where the resulting clusters were used for behavior classification, enabling the identification of distinct charging profiles.

Similarly, k-means clustering was utilized in [52] to identify patterns in EV charging profiles across different UK counties, providing a meaningful segmentation of charging behaviors. Hierarchical clustering has also been effective in revealing distinct groups, as shown in [53], where clusters representing different types of chargers were identified.

GMMs and BMMs offer a probabilistic approach to clustering, allowing for more nuanced representations of charging behaviors. GMMs have been used to uncover clusters representing different charging scenarios in [54], while BMMs were applied in [55] to capture the underlying structure of charging profiles.

DBSCAN clustering has revealed distinct clusters representing different charging scenarios, providing valuable insights into the diverse patterns exhibited by EV users [56]. GMMs were utilized in [57] to generate EV profiles based on charging events, capturing the underlying distribution of the data.

Performance evaluation of clustering techniques varied across studies. Some studies employed non-parametric statistical estimation techniques such as Kernel Density Estimation (KDE) and Dirichlet Kernel Density Estimation (DKDE) for predicting session duration and energy consumption [58] and [59]. The DKDE method demonstrated superior performance compared to GKDE in these cases.

In pursuit of enhanced prediction accuracy, hybrid estimators combining different estimation techniques have been proposed. For example, a hybrid estimator combining GKDE and DKDE was proposed in [60], leveraging the strengths of both methods to provide more accurate predictions of charging session duration and energy consumption.

Overall, clustering techniques have emerged as powerful tools for understanding EV charging behavior. They allow for the discovery of meaningful patterns and groupings within charging data, enabling researchers to develop more effective strategies for managing charging infrastructure, optimizing resource allocation, and enhancing the efficiency and sustainability of EV charging systems.

In the realm of predictive modeling for EV charging behavior, the existing literature primarily focuses on specific aspects such as feature selection, regression models, and clustering techniques. However, there is a lack of comprehensive approaches that consider multiple factors, such as incorporating additional features from historical behavior, exploring the dependence of charging behavior on various variables, and modeling both individual and aggregated EV behavior. This study aims to address this gap by adopting a holistic approach that combines these elements. By extracting additional features, exploring charging behavior dependence, and modeling individual and aggregated behavior using supervised machine learning, this research seeks to enhance the accuracy and understanding of EV charging predictions.

2.4 EV Charging Scheduling Optimization

In the field of EV charging scheduling optimization, various approaches and methodologies have been employed to develop strategies and algorithms for efficient charging management. Mathematical programming models, such as quadratic optimization, dynamic programming, mixed-integer linear programming (MILP), and linear programming, are commonly used in this area [61], [62].

Quadratic optimization and dynamic programming have been successful in reducing customer energy bills and mitigating peak load penalties, particularly for large bus fleets [63]. These methods enable the careful management of charging schedules, leading to cost savings and improved load balancing.

To address the impact of EV charging on residential distribution grids, MILP approaches have been employed. Studies have investigated the modeling of the neutral conductor in unbalanced Low Voltage Distribution Networks (LVDNs) and emphasized its importance in obtaining realistic values for parameters like voltage profile and power loss [64]. Additionally, MILP-based approaches have been proposed to minimize daily charging costs by considering electricity prices and charging station availability [65].

Decentralized charging methods have also been proposed, allowing drivers to locally select their charging programs. This approach reduces communication costs and computational complexity, empowering drivers to make charging decisions at a local level and improving the overall efficiency and convenience of EV charging [66].

Optimization techniques have been applied to determine optimal charging station placement, considering cost-effectiveness and strategic selection. Linear programming methods have been utilized to identify advantageous placement strategies, considering multiple objectives and constraints [67], [68]. In one study, a mixed-integer programming-based optimization strategy was developed with forecast exaggerations of the day-ahead scheduling, aiming to minimize electricity costs for EV users while managing peak demand for grid operators. The strategy demonstrated improvements in the load factor of the local distribution network, peak-to-average ratio, and cost reductions [69].

Integrating renewable energy sources and addressing multi-objective optimization in EV charging scheduling have also received attention. Researchers have explored using EV batteries as energy storage systems in charging stations for price arbitrage and renewable power integration. This problem is often formulated as a dynamic program, and optimal scheduling strategies have been discussed [67]. A real-time multi-objective optimization model was developed to simultaneously reduce electricity costs, minimize battery degradation, and level grid stress while meeting EV users' departure time demand [70]. The proposed model demonstrated its capability for real-time charging and discharging scheduling, considering multiple objectives. Another study focused on finding an optimal solution to minimize load while meeting users' satisfaction using data collected from different types of charging stations, with residential and office charging sites showing the greatest potential for load reduction and minimal customer impact [71].

In the context of EV aggregator planning, probability frameworks and stochastic optimization techniques have been employed. A probability framework was developed for the optimal planning of EV aggregators, leveraging stochastic optimization to efficiently coordinate EV charging schedules and contribute to system optimization [72].

The Smart Microgrid (SMG) model has been developed to incorporate stochastic scheduling of EVs and price volatility in energy markets. This model takes into account uncertainties related to EV behavior and uses K-means clustering techniques to optimize charging schedules and integrate renewable energy sources [73].

Game theory-based approaches have emerged as another valuable approach in EV charging scheduling optimization. By modeling the charging process as a non-cooperative game, these methods aim to find equilibrium charging schedules that maximize individual user satisfaction while considering system-level constraints.

In this approach, the focus is on finding suitable solutions for the involved players rather than solely optimizing costs, which is typical in optimization-based methods. Game theory-based approaches have been implemented to achieve this goal and have shown promising results. For instance, studies have demonstrated significant reductions in travel time and road traffic density, leading to energy savings [74]. Cooperative game theory was applied in [75] to investigate the formation of "coalitions" between employers and employees for scheduling the charging and discharging of EVs. The findings indicated that such scheduling can effectively reduce annual power costs for both parties, highlighting the potential benefits of cooperative strategies.

One application of game theory in EV charging scheduling is to consider it as a non-cooperative game where each aggregator determines the start time and energy profile for charging EVs to minimize the total cost of charging energy [76]. By employing such strategies, energy savings can be achieved.

Stackelberg game modeling has also been proposed to address the pricing mechanisms of aggregators and charging EVs [77]. This method involves defining upper and lower levels for pricing, where the upper level represents the price of grid electricity, and the lower level encompasses the pricing mechanism set by the aggregator. By maximizing the profit of the aggregator while minimizing the charging cost for each EV, taking into account the pricing factors, game theory and Nash equilibrium calculations can be employed.

Another approach is to use non-cooperative optimization methods based on matching algorithms for EV charging scheduling [78]. The authors aimed to balance the utilization ratio between charging stations and EVs by employing a matching algorithm. They argue that the Stackelberg equilibrium outperforms the Nash equilibrium in considering the dynamics of the system.

Metaheuristic algorithms have gained significant attention in the field of EV charging scheduling optimization. These algorithms, such as genetic algorithms, particle swarm optimization, and simulated annealing, provide efficient and scalable solutions to handle the high computational complexity of large-scale optimization problems.

In a smart grid environment, a hybrid genetic algorithm was proposed to optimize EV charging schedules by considering factors such as user mobility patterns, electricity prices, and grid constraints [79]. This approach demonstrated superiority over conventional methods, showing the effectiveness of metaheuristic algorithms.

Other metaheuristic algorithms, such as Grey Wolf Optimization (GWO) and Improved Binary Grey Wolf Optimization (IBGWO), have been utilized to optimize real-time charging using energy storage and photovoltaic systems, resulting in cost reduction and improved system performance [80]. Centralized Genetic Algorithms (GA) have been employed to find optimal charging strategies by considering power prices and battery characteristics [81]. These algorithms consider different objectives and constraints to achieve efficient charging schedules.

To address the placement of charging stations, metaheuristic algorithms such as genetic algorithms have been utilized. A GA-based optimization model was proposed to determine the number and locations of charging stations, resulting in cost reduction and improved service quality [82]. Multi-objective models have also been developed, aiming to maximize traffic flow in traffic networks while minimizing power loss in distribution networks, providing an optimal compromise [83].

Fuzzy control methods have been employed to address uncertainty in price within the upstream grid, providing robust scheduling for EV charging [84]. Particle Swarm Optimization (PSO) has been used to maximize the average state of charge (SoC) of vehicles connected to the grid, effectively handling overloading and optimizing power utilization [85].

Machine learning techniques, such as reinforcement learning, deep reinforcement learning, and artificial neural networks, have also been applied to optimize EV charging schedules by leveraging historical data and predicting system behavior. These techniques have shown promising results in reducing peak load, costs, and waiting times while improving cost reduction and system performance [86]–[91]

In the field of EV charging optimization, several intelligent charging control algorithms have been proposed to enhance the charging process and achieve various objectives. These algorithms actively determine the most appropriate charging station for EV drivers, leading to reduced charging expenses and preventing the overloading of transformers [92]. Furthermore, algorithms have been developed to improve the scheduling of online charging requests based on user preferences and needs, optimizing the overall charging experience [93].

Model Predictive Control (MPC)-based smart charging strategies have emerged as effective approaches to schedule EV charging. These strategies consider the uncertainty of future charging demands and aim to reduce peak electricity demand by implementing optimized charging schedules [94]. Similarly, an intelligent charge scheduling model utilizing a heuristic algorithm has been introduced to minimize EV charging costs in residential and commercial charging stations, providing cost-saving benefits for EV owners [95].

Addressing the potential challenges associated with the increased penetration of EVs, research has focused on utilizing power more efficiently during off-peak hours. By optimizing the charging patterns and load distribution, stress on the grid can be alleviated, ensuring the reliable and efficient operation of the electrical system [96].

The existing literature on EV charging scheduling optimization has primarily focused on specific aspects such as cost minimization, grid stress reduction, and user demand satisfaction. This study seeks to further examine the trade-off between these competing objectives, as well as the implication of using predicted charging behavior to schedule charging sessions.

2.5 Summary

The literature review reveals several important findings and gaps in existing research on EV charging behavior and scheduling optimization. Uncoordinated charging patterns can strain the electric grid, leading to increased costs, grid instability, and limited renewable energy utilization. While previous studies have primarily focused on public charging stations, understanding residential charging behavior is crucial due to its potential impact on residential grids. Furthermore, existing predictive modeling approaches have limitations in incorporating multiple factors, exploring dependence on various variables, and modeling individual and aggregated behavior.

To address these gaps, this study aims to conduct a comprehensive analysis of EV charging behavior by leveraging data from household charging stations. By considering a wide range of charging scenarios and user behaviors, a more comprehensive understanding of EV charging patterns can be achieved. The study also adopts a holistic approach that incorporates additional features, explores charging behavior dependence, and utilizes supervised machine learning to enhance the accuracy of charging predictions. This research contributes to the development of robust predictive models that can inform decision-making for managing charging infrastructure and optimizing grid operations.

3. EV CHARGING BEHAVIOR ANALYTICS

3.1 Overview

Over the past decade, the sale of EVs has grown rapidly in the United States, as reported by the International Energy Agency (IEA). In 2021, approximately 6.6 million EVs were sold, which is more than double the 3 million sold in 2020 and more than triple the 2.2 million sold in 2019. This marks a major, positive change to the automotive industry, which is one of the largest economic sectors in the U.S., supporting millions of American jobs. As of 2019, the U.S. had at least 250,000 jobs in the manufacture, sale, and maintenance of electric vehicles, and this number continues to grow as EV production and sales expand [97]. However, the swift progress of electric vehicles (EVs) has presented obstacles to both power grids and transportation networks. To ensure that EVs can interact effectively with electrified transportation networks, it is crucial to accurately understand the usage patterns of EV users. EV charging behavior analytics typically involves the collection and analysis of data from various sources, such as charging station data, EV user data, grid data, and environmental data. This data can be analyzed using various analytical techniques, such as statistical analysis, data visualization, machine learning, and predictive modeling, to extract meaningful information and insights. Some of the key areas that can be analyzed in EV charging behavior analytics include charging patterns (e.g., peak charging times, charging durations), charging preferences (e.g., charging speed, location choices), charging demand (e.g., number of charging sessions, energy consumption, charging power requirements), charging efficiency (e.g., optimization of charging schedules, peak demand management), user behavior and preferences (e.g., charging frequency, user profiles). The insights gained from EV charging behavior analytics can help stakeholders optimize the utilization of charging infrastructure, improve customer satisfaction, enhance charging service offerings, reduce costs, minimize grid impact, and support the growth of the EV market. It can also inform policy and investment decisions related to EV charging infrastructure development, regulation, and incentives.

3.2 EV User Charging Behavior Analysis Framework

This chapter focuses on analyzing EV user charging behavior at both household and public charging stations using real collected data from accessible Level 2 charging ports in

Household: The total dataset has 275,827 charging sessions obtained from 485 household charging stations located in Omaha, NE from Apr 2018 to Dec 2022 as shown in Table 3-1.

Public: The total dataset has 84,577 charging sessions obtained from 160 public charging stations located in Nebraska from Jan 2014 to Dec 2022 as shown in Table 3-2.

Table 3-1: Cumulative summary of the usage of household charging stations.

	No. of cumulative Charging Ports	**No. of Connection Sessions**	**Energy (kWh)**
2018	123	10,487	119.1
2019	231	45,921	547.9
2020	360	48,022	611.6
2021	477	82,898	1,137
2022	485	88,499	1,244
Total	**485**	**275,827**	**3,660**

Table 3-2: Cumulative summary of the usage of public charging stations.

Year	**Cumulative No. of Stations**	**Cumulative No. of Users**	**No. of Sessions**	**Energy (MWh)**
2014	7	45	1,621	8,350
2015	15	97	1,962	14,114
2016	34	211	2,825	23,871
2017	40	427	4,361	34,715
2018	47	756	7,148	61,136
2019	58	1,137	9,471	108,238
2020	83	1,250	7,228	88,426
2021	136	3,530	17,086	210,054
2022	160	5,678	32875	473,886
Total	**160**	**5678**	**84,577**	**1,022,790**

For each session, the following information is considered: session start and end time (including day of week and time of day), connection duration - the duration between the beginning and the end of the session, charging duration- the time required to charge the EV fully or partially, kWh consumed during the charging session, and unique driver ID. Figure 3-1 shows the framework used in this section.

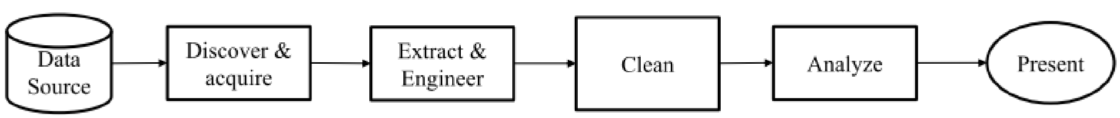

Figure 3-1: EV charging analysis framework.

3.2.1 EV Connection Start Time

The start time of a charging session is simply the exact time that a user connects their vehicle. It is a critical variable in charging behavior analysis, as it helps to determine the most common times for charging sessions in an area. This analysis can be utilized for grid planning to avoid potential transformer overload from excessive unanticipated simultaneous charging.

Figure 3-2 shows the distribution of start times, in hourly increments, for the sessions in household charging stations data. Starting at 6-7 am, each hour shows a mostly increasing likelihood of sessions beginning. Around 2-3 pm, the likelihood begins rapidly increasing each hour, up to a peak at 5-7 pm. This can be reasonably assumed to correspond to the hours that many users arrive home from work and/or errands. The number of sessions starting each hour then declines for the rest of the night, except for a spike from 10-11 pm, which may be due to a combination of users with late work shifts, evening activities, or simply choosing to plug vehicles in before bed.

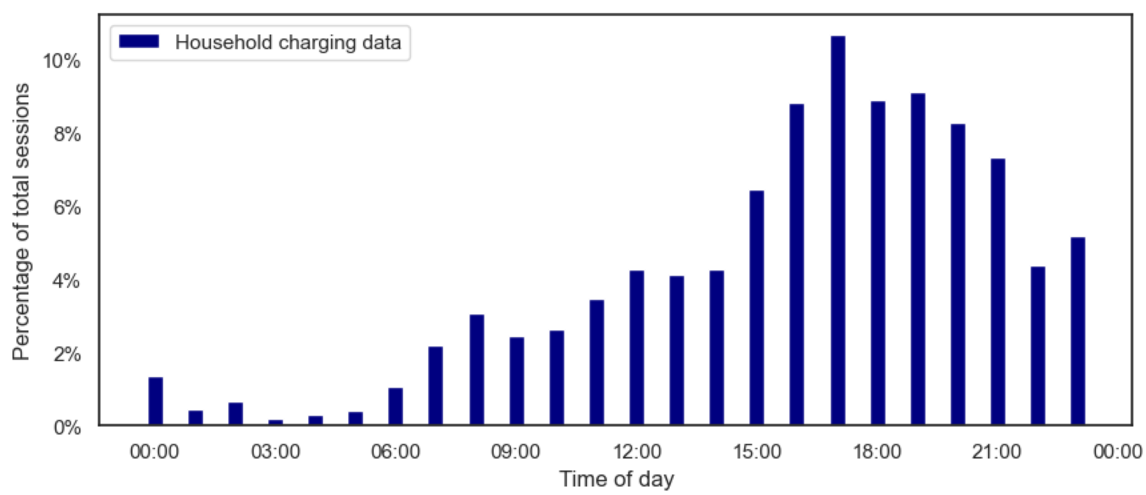

Figure 3-2: The percentage of total household charging sessions with a given start time.

Figure 3-3 shows the distribution of start times, in hourly increments, for the sessions in public charging stations data. Most users use the stations during the day between 6 am and 6 pm. Starting at 6-7 am, each hour shows a rapidly increasing likelihood of sessions beginning with two peaks at 8 am and 1 pm. This can be reasonably assumed to correspond to the hours that many users arrive at their work from home. The number of sessions starting each hour then declines for the rest of the day.

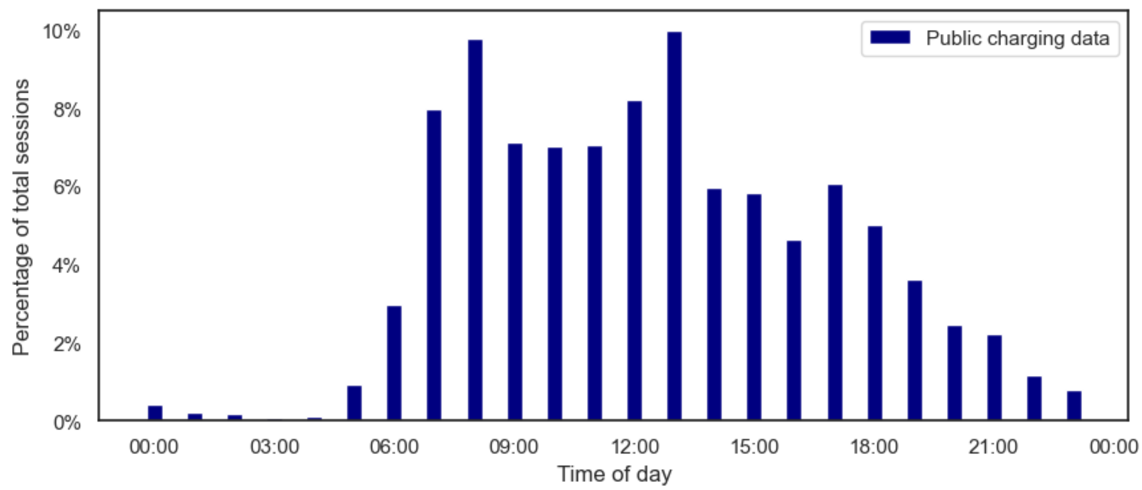

Figure 3-3: The percentage of total public charging sessions with a given start time.

Figure 3-4 and Figure 3-5 analogous to the previous Figures, also show the distribution of connect start, but separate the data by day of the week. Connection start times follow very similar patterns during workdays (typically Monday to Friday) as most EV users have a routine of charging their vehicles during the day while they are at work. This could result in a consistent pattern of EV charging behavior during weekdays, with similar connection start times. On the other hand, weekends (Saturday and Sunday) may exhibit different charging behavior as EV users may be at home or have different usage patterns, leading to variations in connection start times.

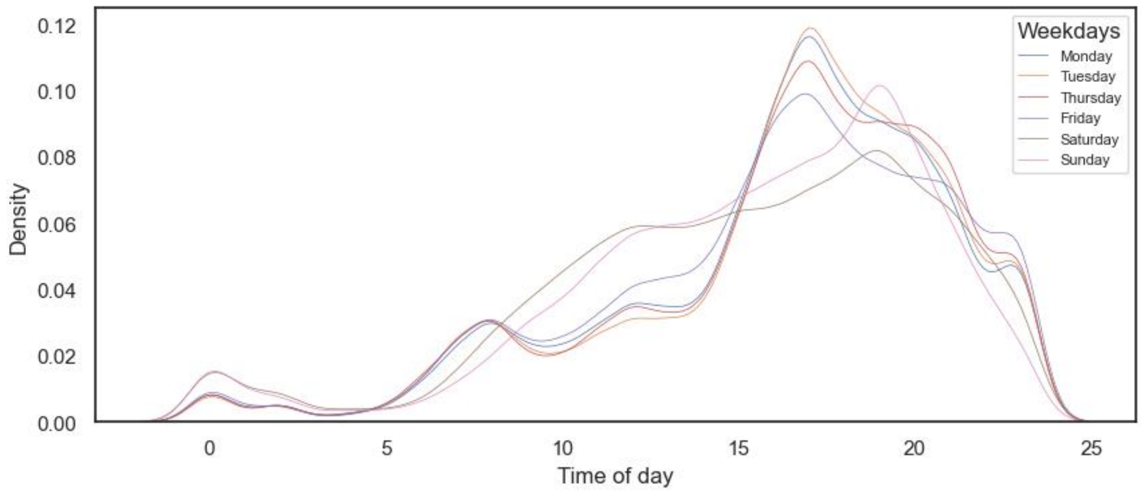

Figure 3-4: The density plot of total household charging sessions with a given start time per day.

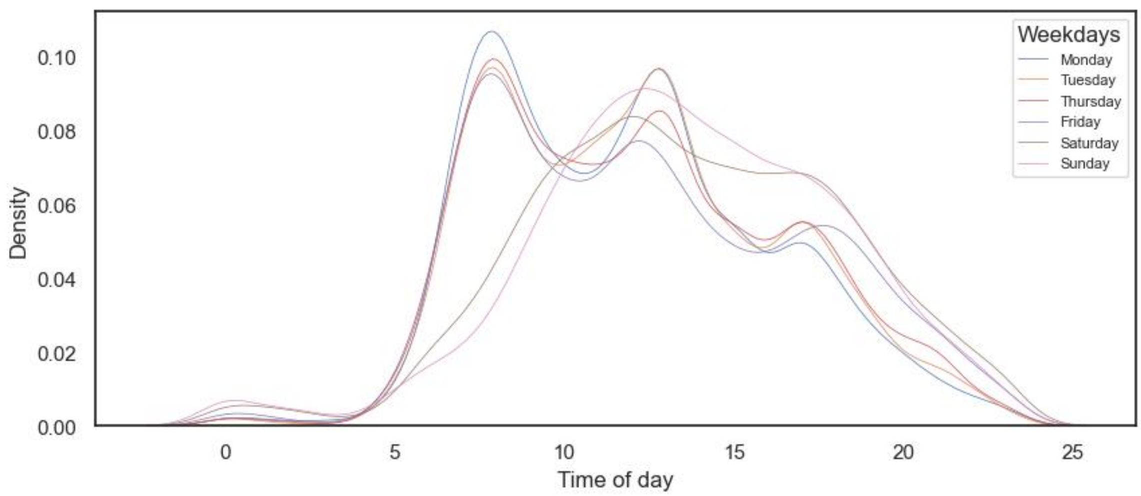

Figure 3-5: The density plot of total public charging sessions with a given start time per day.

3.2.2 EV Connection End Time

The end time of a charging session refers to the time when an electric vehicle charging session or connection is scheduled to end or has ended. It may be used for billing or tracking purposes, as the duration of the charging session is often a factor in determining the cost of the charging service. It is an important factor in effectively planning and managing EC charging sessions.

Figure 3-6 shows the distribution of end times, in hourly increments, for the sessions in household charging stations data. It shows that the plurality of connection end times occurs between 7-8 am, with a rapidly decreasing prevalence in later hours of the day. This can be reasonably assumed to correspond to the hours that users leave for work or other daily activities.

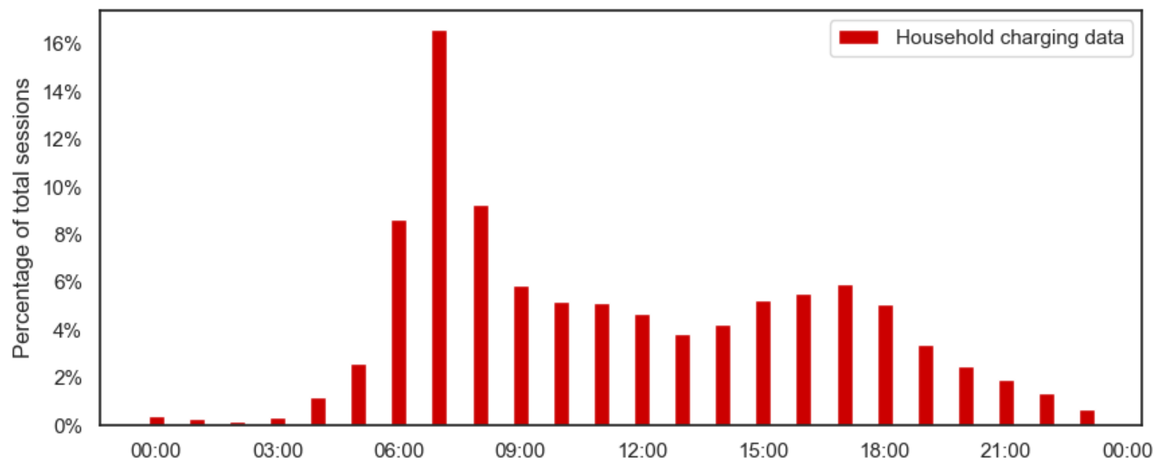

Figure 3-6: The percentage of total household charging sessions with a given start time.

Figure 3-7 shows the distribution of end times, in hourly increments, for the sessions in public charging stations data shows that the plurality of connection end times is much more evenly distributed - times between 7 am and 7 pm are roughly the same with a peak at noon. This can be reasonably assumed to correspond to the hours that many users finish their morning charging sessions.

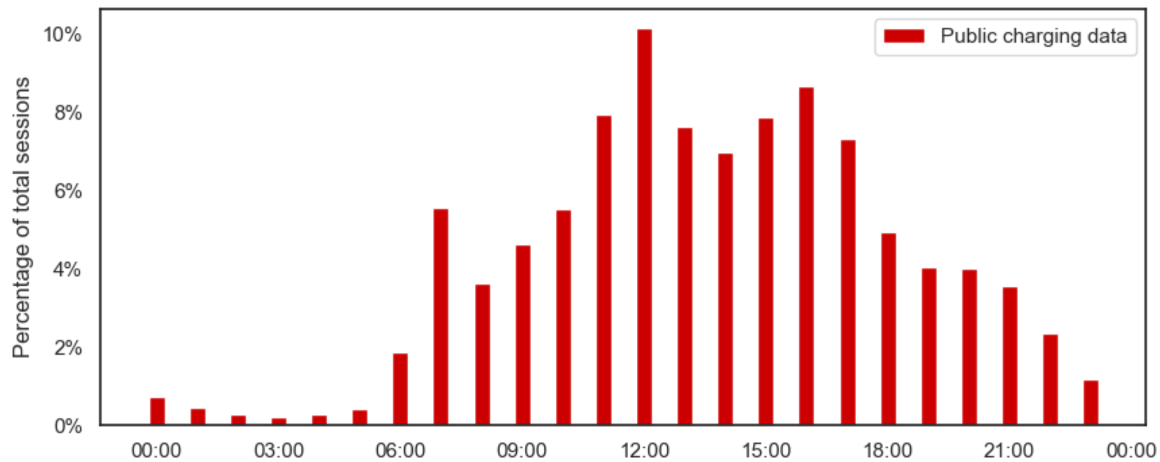

Figure 3-7: The percentage of total public charging sessions with a given end time.

Figure 3-8 and Figure 3-9 analogous to the previous Figures, also show the distribution of disconnect time, but separate the data by day of the week. Connection end times follow very similar patterns during workdays (typically Monday to Friday) as most EV users plug out their vehicles in the morning before heading to their work. This could result in a consistent pattern of EV charging behavior during weekdays, with a similar connection end time. On the other hand, weekends (Saturday and Sunday) may exhibit different charging behavior as EV users may be at home or have different usage patterns, leading to variations in connection end times.

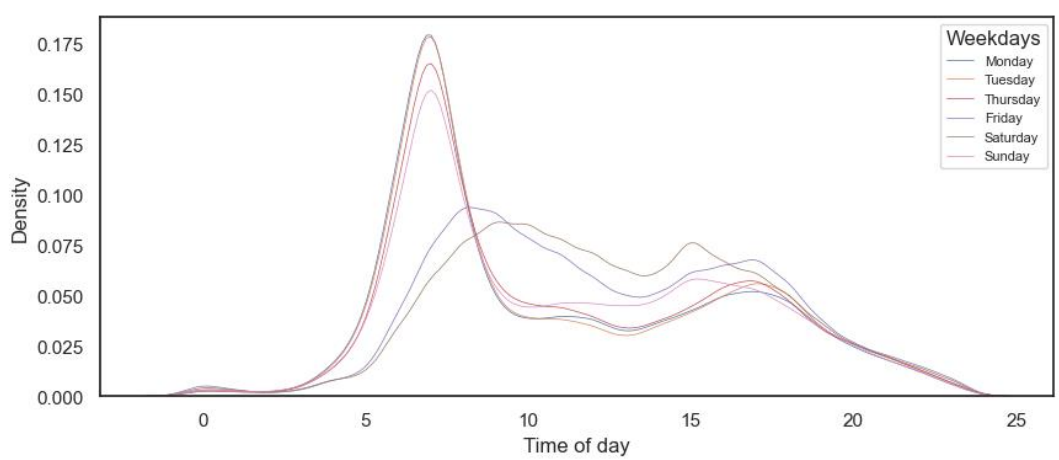

Figure 3-8: The density plot of total household charging sessions with a given end time per day.

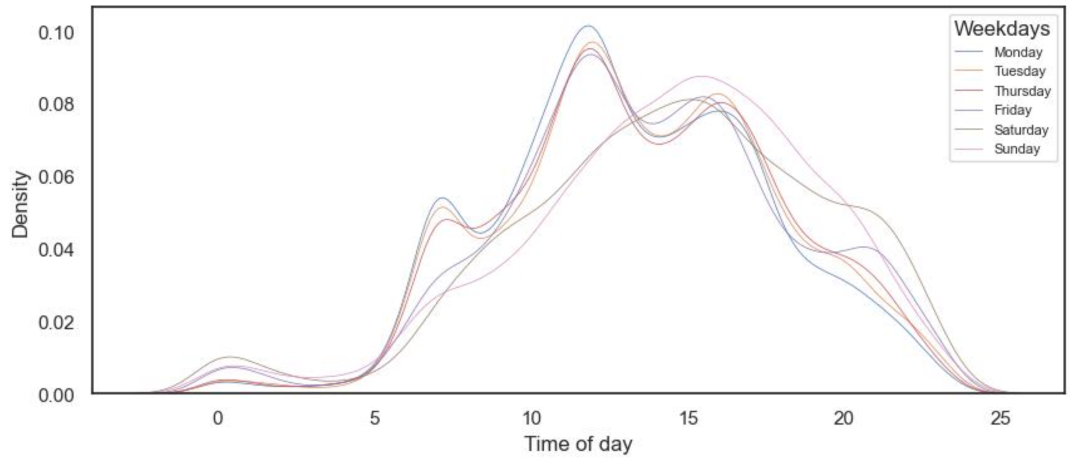

Figure 3-9: The density plot of total public charging sessions with a given end time per day.

3.2.3 Connection Duration

The connection duration for each session is simply the time elapsed between connecting and disconnecting the vehicle from the charging station. It is the duration during which the EV is physically connected to a charging source, either through a charging cable plugged into a charging station or through a wireless charging system, to charge its battery. It can range from a few minutes for a quick top-up to several hours for a complete charge, depending on the charging speed and the battery capacity of the EV. Monitoring and analyzing EV connection duration can provide insights into the charging patterns, charging efficiency, and charging demands of electric vehicles, which can be useful for optimizing charging infrastructure, managing charging costs, and improving the overall user experience of EV users.

Figure 3-10 shows the distribution of connection durations, in 1-hour increments of the household charging data. It is observed that many connection durations, 53%, are 8 hours or more, with an average connection time of 11.6 hours. This is expected as the station is dedicated duration for the user and no other user is waiting to use it.

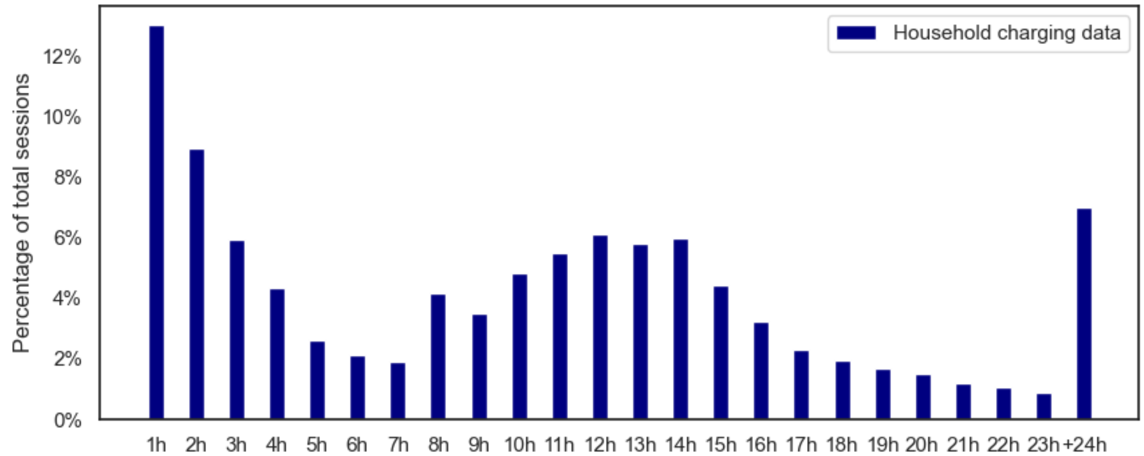

Figure 3-10: The percentage of total household charging sessions with a given connection duration.

Figure 3-11 shows the distribution of connection durations, in 1-hour increments of the household charging data. It is observed that many connection durations, 16.5%, are 8 hours or more, with an average connection time of 6.3 hours.

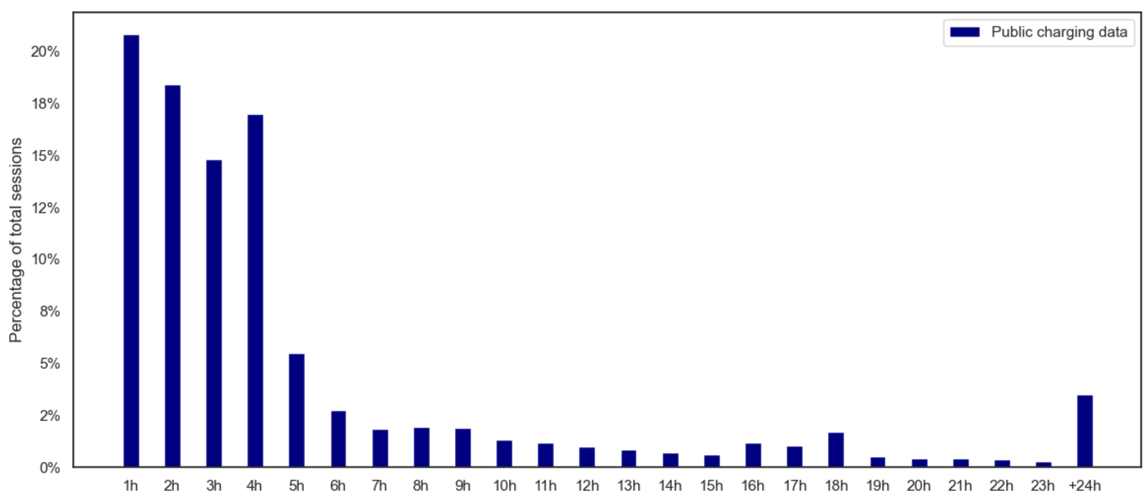

Figure 3-11: The percentage of total public charging sessions with a given connection duration.

3.2.4 Charging Duration

The charging duration for each session refers to the length of time it takes for an electric vehicle (EV) to charge its battery from one state of charge (SOC) to another, typically expressed in minutes or hours. It is the time that an EV spends connected to a charging station or charging point, actively charging its battery to replenish its energy. The charging duration of an EV can vary depending on several factors, including the charging power or rate of the charging station, the battery capacity of the EV, the initial SOC of the battery, and the desired SOC. Higher charging power or rate can result in shorter charging durations, while lower charging power or rate may require longer charging durations. Additionally, the battery capacity of the EV and the desired SOC, such as a partial charge or a full charge, can also impact the charging duration. It is an important parameter to consider when planning EV charging, estimating charging costs, and understanding the charging behavior of EVs. It can be used to optimize the charging process, manage charging infrastructure, and plan for charging requirements, such as charging station availability, charging time, and charging costs.

Figure 3-12 shows the distribution of charging durations for all household charging sessions, average charging duration is 2.4 hours, with only 9.4 % of sessions charging for longer than five hours.

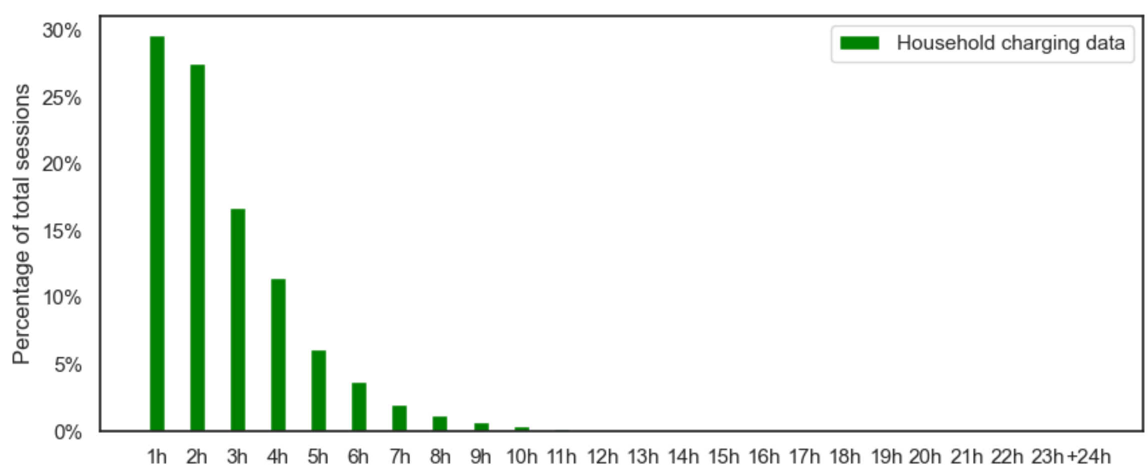

Figure 3-12: The percentage of total household charging sessions with a given charging duration.

Figure 3-13 shows the distribution of charging durations for all public charging sessions, in 1-hour increments. It is observed that many charging durations, 91.2%, are 5 hours or less, with an average connection time of 2.7 hours.

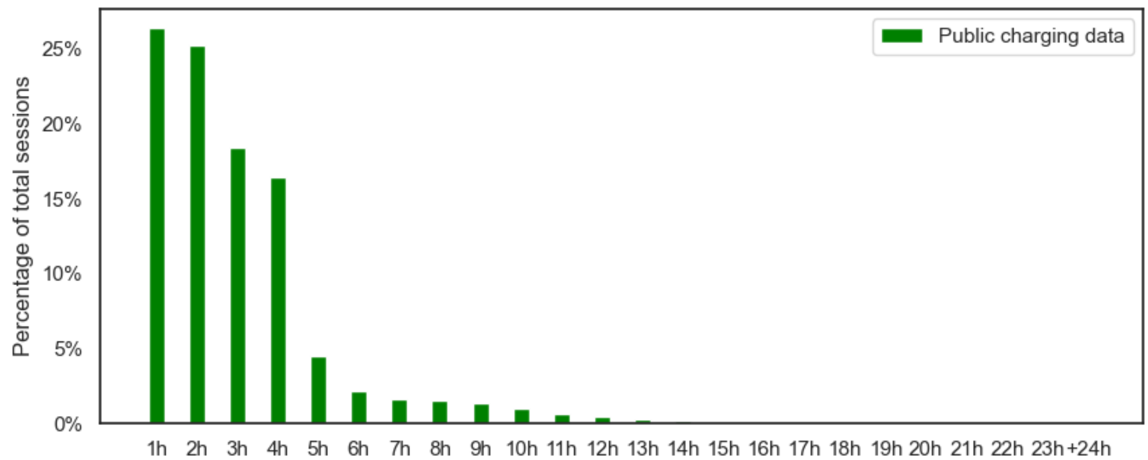

Figure 3-13: The percentage of total public charging sessions with a given charging duration.

3.2.5 Idle Duration

The idle duration for each session refers to the duration during which an electric vehicle (EV) is not actively charging but remains connected to a charging station or charging point without drawing any charging power [98]. In other words, it is the time when an EV is parked and connected to a charging station or charging point, but the charging process is not actively taking place. The duration time can occur for various reasons, such as when an EV has completed charging and remains connected to the charging station or charging point while waiting for the driver to return, when the charging session has been interrupted or paused by the user or the charging infrastructure, or when the EV is parked at a charging station without the intention of charging. It can help in optimizing the operation and management of charging infrastructure, identifying potential issues or inefficiencies in the charging process, and planning for charging requirements, such as charging station availability, parking duration, and charging costs.

The idle duration of each session is calculated by subtracting the charging duration from the connection duration. While roughly 26.3% of sessions have an idle duration of less than one hour, there is a wide distribution of observed times, and the average connection duration is 8.4 hours longer than the charging duration. This idle duration could potentially be used for Vehicle-to-Grid during peak hours, to help the utility manage the demand.

Figure 3-14 shows the distribution of idle durations, in 1-hour increments. It is observed that many idle durations, 41%, are 8 hours or more, with an average idle duration of 8.7 hours.

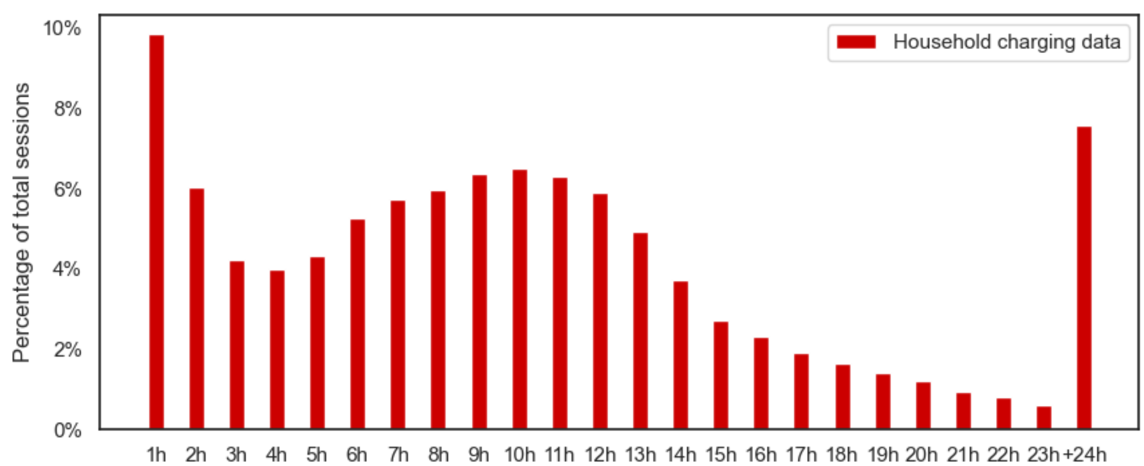

Figure 3-14: The percentage of total household charging sessions with a given idle duration.

Figure 3-15 shows the distribution of idle durations, in 1-hour increments. It is observed that many idle durations, 9.5%, are 8 hours or more, with an average idle duration of 3.6 hours.

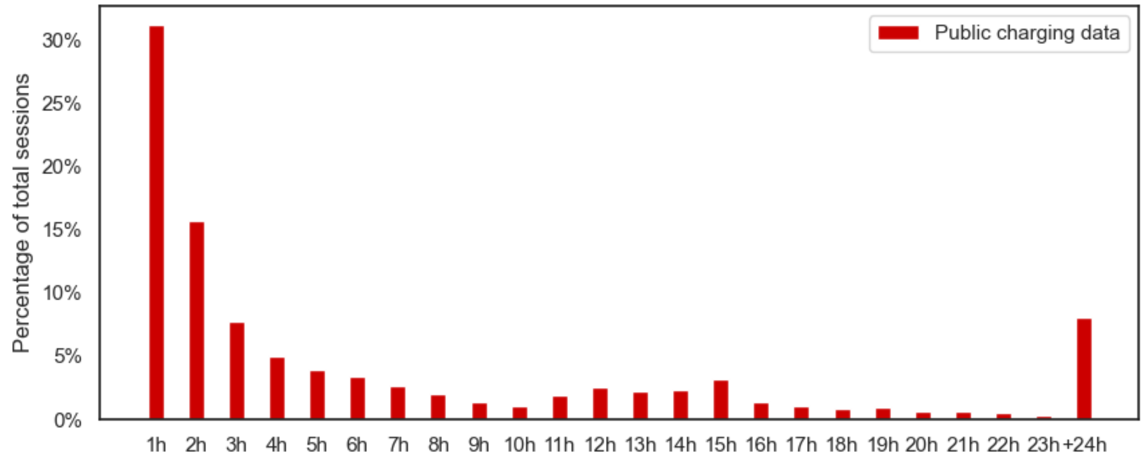

Figure 3-15: The percentage of total public charging sessions with a given idle duration.

3.2.6 Charging Demand (in kWh)

The charging demand for each session refers to the level of charging power or energy required by an electric vehicle (EV) at a given time. It can be influenced by various factors such as the charging power or rate selected by the users, the charging infrastructure capacity, the charging session duration, and the time of day or season. It can also vary based on the charging network, location, and type of charging (e.g., Level 1, Level 2, DC fast-charging). Like the session start and end times, analysis of session energy usage is vital for residential grid planning as EV adoption increases as it helps in optimizing the capacity and distribution of charging stations, predicting future charging demand, identifying peak charging periods, managing charging costs, and ensuring efficient and reliable charging services for EV users.

Figure 3-16 shows the histogram of energy consumed each session, for the full dataset. It is observed that 88% of sessions consumed less than 30 kWh, with an average energy consumption of 14 kWh.

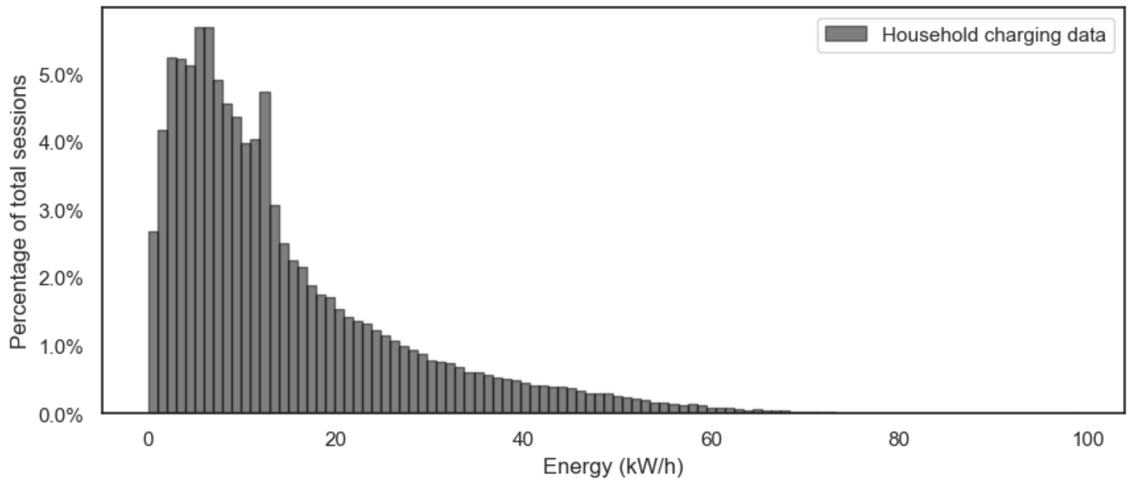

Figure 3-16: Histogram of energy consumed at household charging stations each session.

Figure 3-17 shows the histogram of energy consumed each session, for the full dataset. It is observed that 92% of sessions consumed less than 30 kWh, with an average energy consumption of 12.7 kWh.

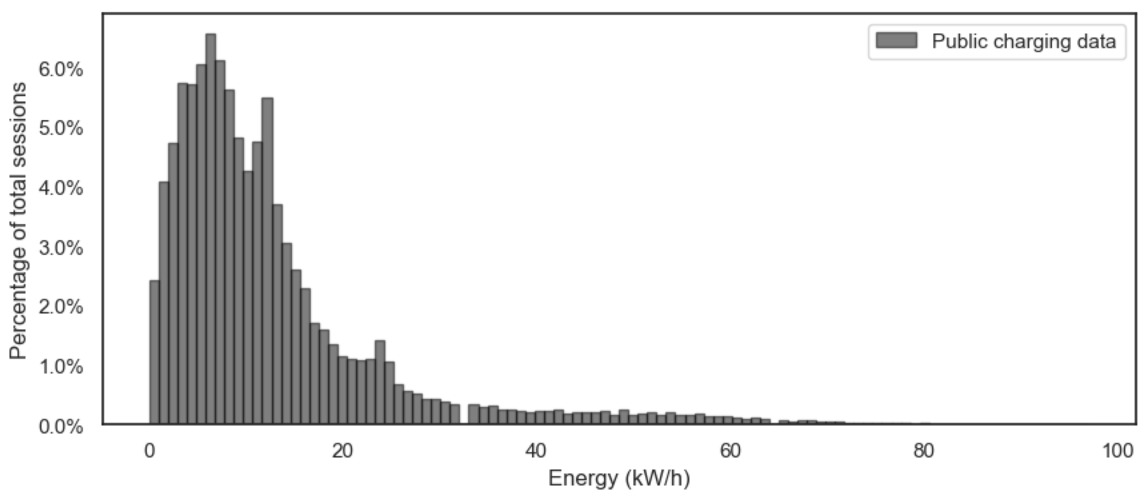

Figure 3-17: Histogram of energy consumed by public charging stations each session.

3.3 Conclusions

In conclusion, the analysis of EV charging session start times, end times, connection durations, and energy consumption provides valuable insights into user behavior and charging patterns. As shown in Table 3-3, The findings indicate distinct patterns in start times, with household charging peaking in the evening and public charging showing concentrated usage during the daytime. Weekdays exhibit consistent charging routines, while weekends exhibit variations reflecting different behaviors and routines. The majority of charging sessions end in the morning for households, while public charging stations display a more even distribution of end times.

The analysis of connection durations reveals that a significant proportion of connections, both in households and public charging, last 8 hours or more. This suggests dedicated usage of charging stations without waiting time. Additionally, the average charging durations are relatively short, with most sessions lasting under five hours. The analysis of energy consumption demonstrates that the majority of charging sessions consume less than 30 kWh, indicating relatively low energy requirements. These findings have implications for infrastructure planning and energy management, providing insights into typical energy consumption and helping optimize charging infrastructure to meet user needs.

Overall, using a single city as a case study, this chapter contributes to a better understanding of EV charging behavior and patterns, highlighting the importance of considering different variables such as start times, end times, connection durations, and energy consumption. There exist clear patterns in user behavior for the starting time of sessions, as well as the connection duration, charging duration, and subsequent energy demand. The weekday behavior is highly consistent for each day of the week, and relatively small differences exist between Saturday and Sunday charging behavior. The existence of such patterns is evidence that charging behavior may be predictable with some accuracy.

The existence of such patterns is evidence that charging behavior may be predictable with some accuracy.

Table 3-3: EV charging analysis results.

	Household Charging Data		**Public Charging Data**	
	Peak	**Average**	**Peak**	**Average**
EV Connection Start Time	5-7 pm		6-7 am 12-1 pm	
EV Connection End Time	7-8 am		12-1 pm 4-5 pm	
Connection Duration		11.6h		6.3h
Charging Duration		2.4h		2.7h
Idle Duration		8.7h		3.6h
Charging Demand		14 kWh		12.7 kWh

4. EV CHARGING SCHEDULING

4.1 Overview

In the previous chapter, an extensive analysis of EV charging behavior was conducted, revealing valuable insights into the typical charging patterns observed among EV users. This analysis not only provides a deep understanding of EV charging behavior but also sheds light on its implications for the planning, design, management, and optimization of EV charging infrastructure. It also helps stakeholders gain valuable insights into EV user preferences, demand patterns, and market trends, enabling them to effectively meet the increasing electrical demand resulting from the widespread adoption of EVs and integrate them into smart grid systems.

Building upon these findings, the subsequent chapter aims to delve deeper into the relationship between the analyzed charging behavior and scheduling. The central focus will be to investigate how the insights derived from the analysis of typical charging behavior can influence and inform the development of a simple scheduling routine.

This chapter delves into the potential impacts and challenges associated with scheduling EV charging sessions by harnessing the understanding of typical charging behaviors. It specifically focuses on addressing decision-making processes related to cost minimization and peak load reduction. Furthermore, it investigates the advantages of controlled charging over uncontrolled charging, particularly in light of the increasing penetration of EVs. The chapter undertakes a comprehensive exploration of various EV charging scheduling techniques, utilizing the knowledge acquired from the analysis. Theoretical frameworks and practical examples are presented to provide a holistic understanding of the opportunities and solutions that emerge at the intersection of charging behavior analysis and scheduling. The ultimate goal is to unlock the vast potential of EV charging scheduling to shape a future of transportation that prioritizes sustainability and environmental consciousness.

4.2 EV Charging Scheduling Framework

In this case study, a single day of charging is optimized using a subset of sessions, with a distribution of start times and end times representative of the dataset. Specifically, it is assumed that for each user, on each day the following information is known:

$TS_i =$ time the user connects to the charging station.

$TE_i =$ time the user disconnects from the charging station.

$Ch_i =$ charging duration for the user.

The time resolution at which to calculate energy demand and generate the charging schedule is somewhat flexible. Higher resolutions will more accurately control instantaneous power demand, at the cost of computational efficiency. The following experiments are performed at a resolution of one minute.

One objective of optimization is the minimization of peak demand for the system. This peak demand is not simply the peak EV load, but the sum of the EV load and the existing residential load. The size of the system is both important and flexible – optimization can be performed for a small number of EV users and households, or a large-scale system of both. In this case study, the existing residential demand profile is taken from a publicly available dataset of 114 single-family apartments [99]. Figure 4-1 shows the aggregated load profile for a single day, February 15th, 2016.

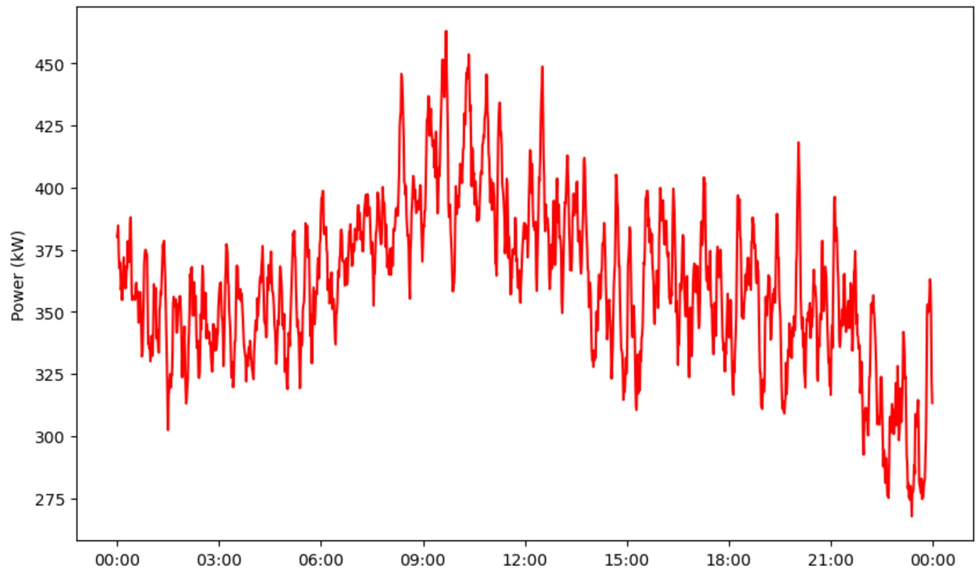

Figure 4-1: Aggregate demand for 114 single-family apartments on a weekday in February 2016 [99].

The second optimization objective is the minimization of total electricity cost for EV charging. The cost per kWh of electricity can vary throughout the day, and a successful optimization of EV charging will shift more of this load to hours with less expensive electricity. Figure 4-2 shows an electricity cost profile for a single weekday in February, obtained from the Nebraska Public Power District website (NPPD) [100]. The daily cost of charging is calculated as the product of this cost profile and the total EV demand, summed over all users and time slots.

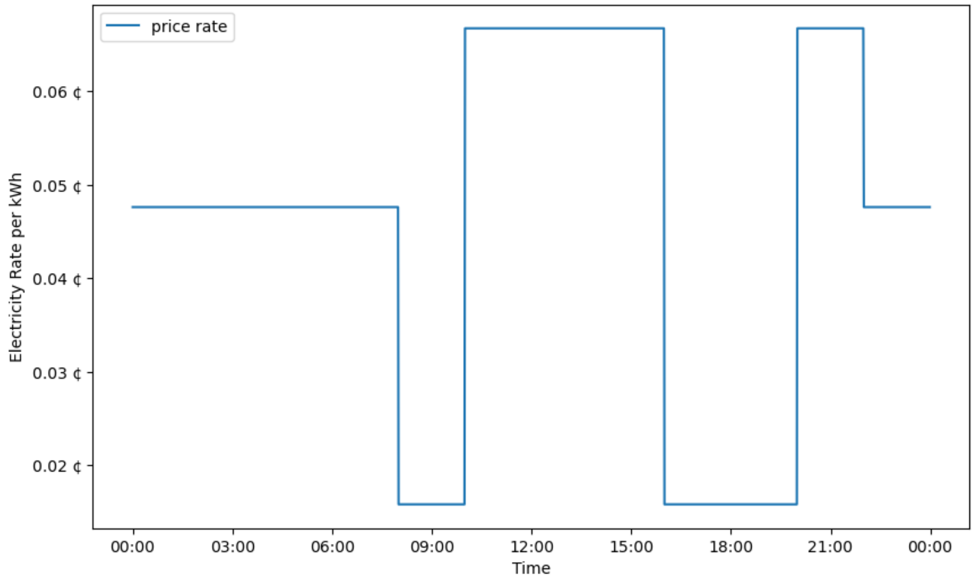

Figure 4-2: Electricity price rate for a weekday in February [100].

What follows is a framework for a single-day optimization scheme for minimizing peak demand and electricity cost, while meeting 100% of each user's charging demand between a given connection and disconnection time. Charging periods are allowed to begin at night and continue into the morning of this same day, but for simplicity, the following equations reflect the standard case where $\mathrm{TE_i} > \mathrm{TS_i}$. Charging periods are forced to be contiguous, to minimize degradation of EV batteries [101], [102]. The charging demand of each EV can be modeled by various profiles, but in the following example is assumed to be constant over the charging duration.

The optimization model must assign a charging period to each vehicle that fits within the user's connection window and meets the user's charging demand. Given these constraints, the optimization model seeks to minimize the peak load and total electricity cost for EV charging for each day. This is formally shown through the following mathematical model. Let $N = \{1,2, \dots, n\}$ denote the set with all users and $M = \{1,2, \dots, m\}$ define the set with all time slots in a day.

4.2.1 Mathematical Model

A charging scheduling framework is applied for a single day of electric vehicle charging, tested on real charging behavior from measured data using various weights assigned to the minimization of electricity cost and peak demand as shown in Figure 4-3

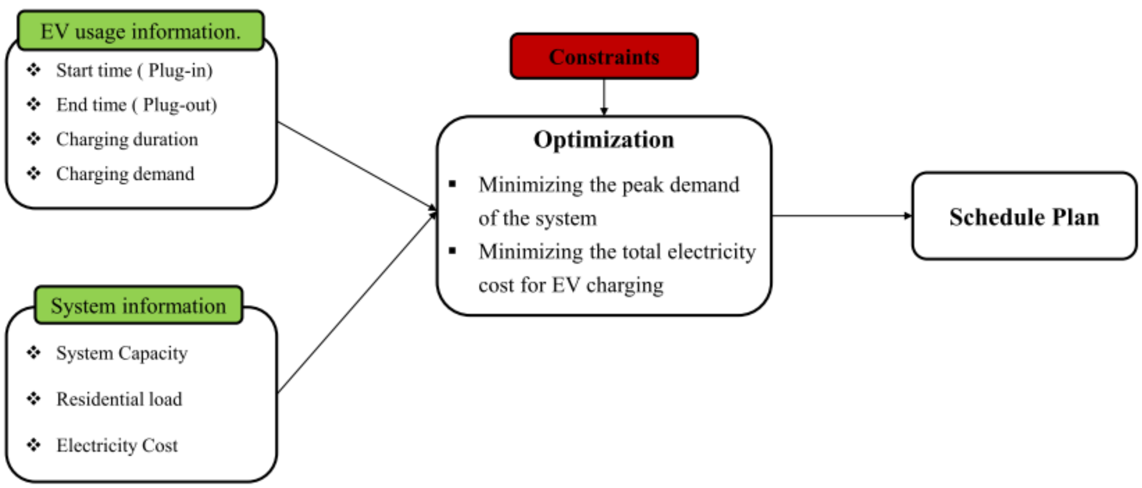

Figure 4-3: EV charging scheduling framework.

- **Parameters**

EC = One-minute energy demand of a charging EV

TC = Total energy capacity of the system for one minute

$Eprice_j$ = Energy price at time slot $j \in M$

D_j = Residential demand at time slot $j \in M$

$Rcost = \sum_{j \in M} D_j Eprice_j$ = Total cost of residential demand

$Rpeak = \max(D_j)\ \forall j \in M$ = Peak residential demand

- **Decision Variables**

$X_{ij} = 1$ if user $i \in N$ charges their vehicle at time slot $j \in M$; 0 otherwise

$Y_{ij} = 1$ if user $i \in N$ starts charging at time slot $j \in M$; 0 otherwise

$w \in \mathbb{R}_+$ = maximum peak workload of the total system

- **Objective Function**

The overall goal of this optimization is to minimize both peak demand and electricity costs. As these are goals with different units, each is normalized relative to the baseline residential peak demand and cost, respectively. The weighting of each objective is controlled by the variables.

$$\min z = Wcost\left(\frac{EC\sum_{i\in N}\sum_{j\in M} X_{ij}Eprice_j}{Rcost} + 1\right) + Wpeak\left(\frac{w}{Rpeak}\right)$$

- **Constraints**

Several constraints must be defined to extract from the optimization process only technically feasible results.

1. Charging demand must be met for each user:

$$\sum_{j\in M} X_{ij} = Ch_i \ \forall i \in N$$

2. Scheduling window should be between start and end times for each user:

$$\sum_{j \in M : j < TS_i, j > TE_i} X_{ij} = 0 \;\; \forall\, i \in N$$

3. System capacity (e.g. transformer) must not be exceeded:

$$\sum_{i \in N} ECX_{ij} \leq TC \;\; \forall\, j \in M$$

4. Determine the maximum peak workload w:

$$D_j + \sum_{i \in N} ECX_{ij} \leq w \;\; \forall\, j \in M$$

5. Vehicles must be charged for consecutive periods:

$$\sum_{j \in M} Y_{ij} = 1 \;\; \forall\, i \in N$$

$$\sum_{j=k}^{k+Ch_i-1} X_{ij} \geq Ch_i Y_{ik} \;\; \forall\, i \in N, k \in \{1, \dots, m - Ch_i + 1\}$$

This mathematical model has been implemented in Python using DOcplex, the IBM Decision Optimization CPLEX Modeling for Python, and solved using CPLEX, a high-performance mathematical programming solver, version 20.1. The study was performed on an AMD A6-3400M APU @1.40 GHz. Because solving this problem using exact methods can be computationally intense, CPLEX was run until an optimality gap of less than 4% is achieved. Several constraints must be defined to extract from the optimization process only technically feasible results.

4.3 Results and Discussions

The following results assume an existing residential load profile as shown in Figure 4-1, and a daily electricity price rate shown in Figure 4-2. The total electricity cost for this day is calculated to be $399, and the peak demand is 463 kW.

A subset of 100 charging sessions is carefully selected. This section aims to ensure that these 100 sessions exhibit a distribution of start times, end times, and charging durations that are representative of the entire dataset. This selection involves considering the various charging patterns observed in the full dataset. The sessions are chosen in a manner that captures the diversity and range of charging behaviors exhibited by the electric vehicle users in the study.

To assess the characteristics of the selected 100 sessions, a comparative analysis is conducted against the full dataset. This analysis helps evaluate how well the subset represents the overall charging behavior observed in the larger dataset. Factors such as the distribution of start times, end times, and charging durations are examined to ensure that the selected sessions align with the patterns observed in the complete dataset.

The meticulous selection of this subset of 100 sessions ensures that it accurately represents the charging behavior observed in the larger dataset. This carefully curated sample serves as a dependable foundation for drawing conclusions and making meaningful inferences about the broader charging patterns and behaviors of the electric vehicle users included in the study.

Each vehicle is assumed to draw a constant 7.5 kW over the full charging duration. Figure 4-4 shows the total demand from EV charging over the course of the day in the uncontrolled scenario, where all vehicles begin charging immediately upon plugging in.

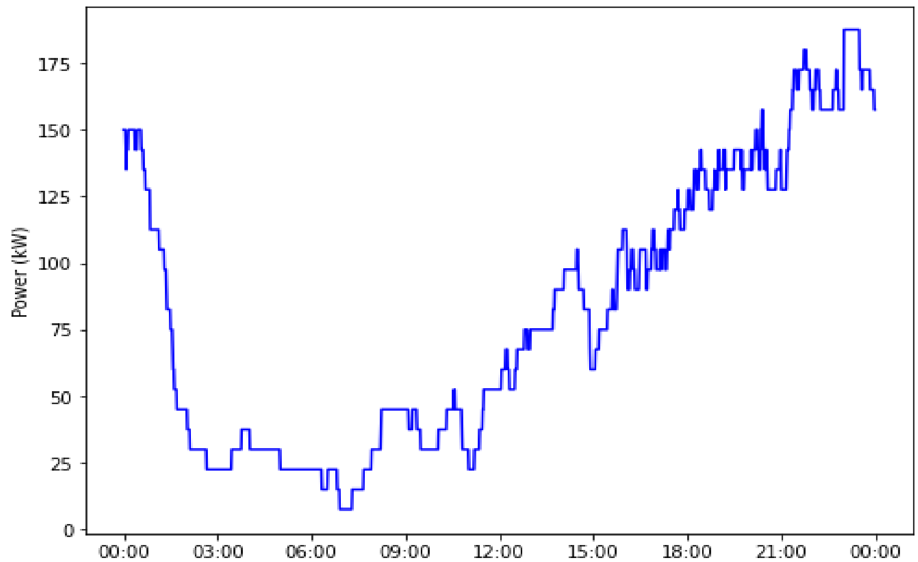

Figure 4-4: Total uncontrolled EV charging demand for 100 sessions over the course of one day.

The total demand in the uncontrolled scenario is the sum of the uncontrolled charging demand and the existing residential load. As the residential load represents 114 single-family apartments, 100 daily charging sessions correspond to a very high degree of EV penetration. In this scenario, the EV load makes up a significant percentage of the total load. Figure 4-5 shows the base residential load in red, and the total load including uncontrolled charging in blue. The uncontrolled charging adds $88 in electricity cost, and raises the peak load to 554 kW, an increase of 91 kW or 20% over the base residential load.

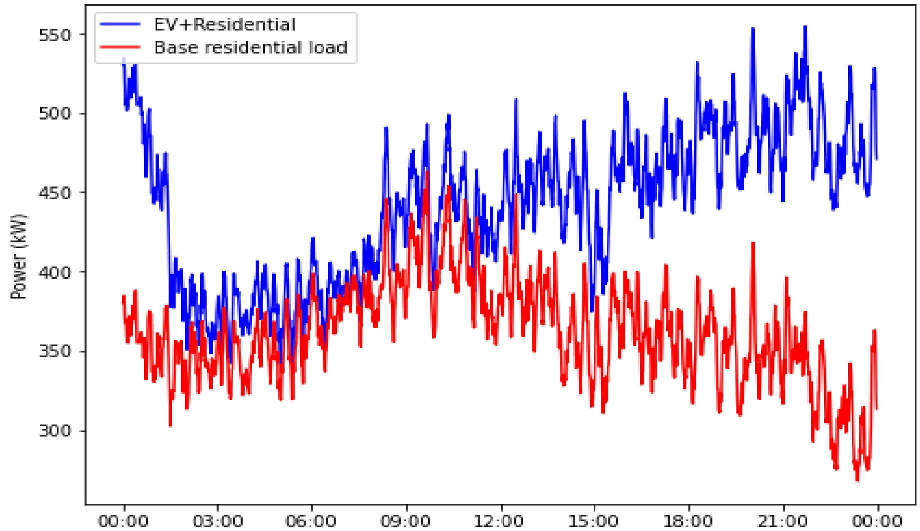

Figure 4-5: Base residential load (red) and total uncontrolled load (blue).

The optimization model is first applied to this scenario with equal weighting for cost and peak minimization. This means that compared to the base residential load, a certain percentage increase in total electricity cost is weighted equivalently to the same percentage increase in total peak demand. The resulting total EV demand for the controlled scenario is shown in Figure 4-6, along with the uncontrolled EV demand for comparison.

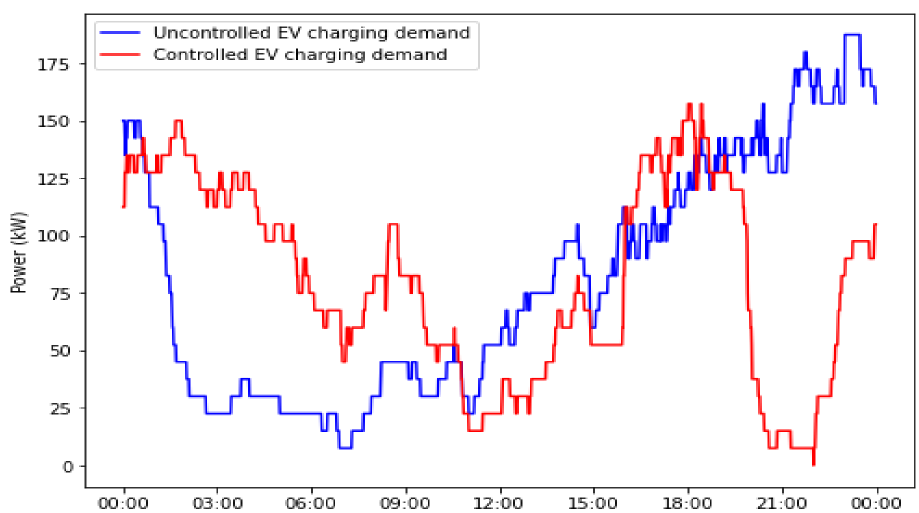

Figure 4-6: Total controlled EV charging demand (red) for 100 sessions, with equal cost and peak demand weights. Uncontrolled EV charging demand is shown in blue.

The effect of the controlled charging on the full residential system in shown is Figure 4-7, with the uncontrolled load in red, and the controlled load in blue. The controlled charging adds $77 in electricity cost, and raises peak demand by 54 kW. Compared to the uncontrolled scenario, the scheduling model has saved $11, representing a 13% savings in EV charging costs. The peak demand is 37 kW lower than the uncontrolled scenario, representing a 7% reduction in peak demand for the full system.

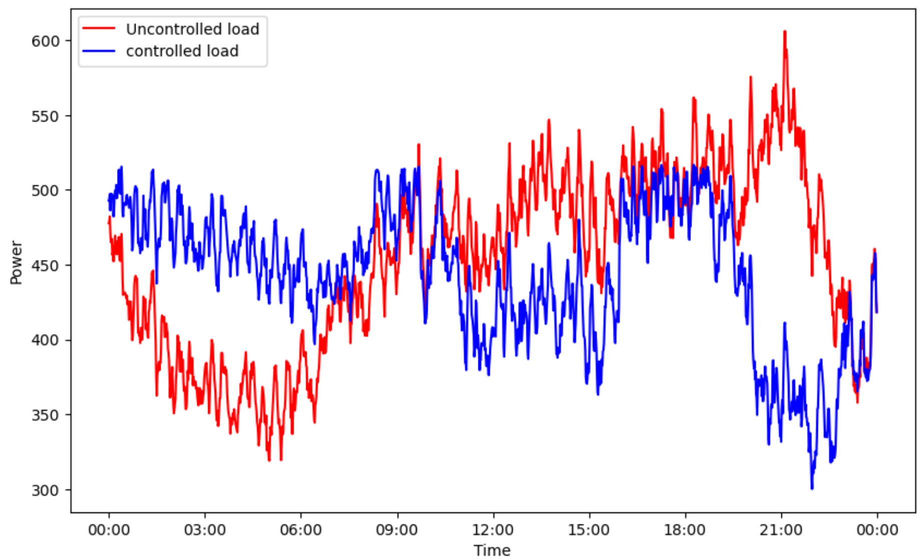

Figure 4-7: Uncontrolled load (red) and controlled load (blue) for equal cost and peak demand weights.

The optimization framework can assign different weights to cost minimization and peak load minimization. As the relative weight of one objective increases, the potential improvement of that metric may increase as well, at the cost of the other. Figure 4-8 shows the results for a scenario in which only cost is minimized. The resulting schedule has a total EV charging cost of $69, $8 less than the equal-weighting schedule. The system peak demand is 663 kW however, 146 kW higher than the equal-weighting schedule.

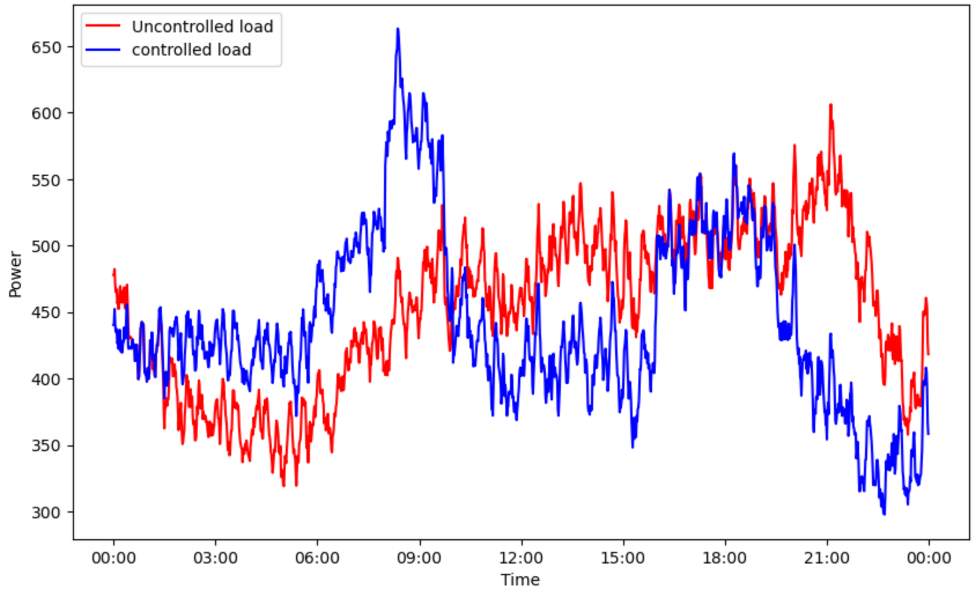

Figure 4-8: Uncontrolled load (red) and controlled load (blue) for cost optimization only.

Conversely, Figure 4-9 shows the results of the model when cost minimization is ignored in favor of peak load minimization. Such a model may be increasingly relevant as EV scheduling becomes more prevalent, as price curves may naturally adjust to reflect flatter aggregate demand profiles. This schedule achieves a peak demand of 489 kW, which is 28 kW lower than the equal-weighting schedule. The total cost of EV charging is $89, however – which is not only $12 higher than the equal-weighting schedule, but $1 higher than the uncontrolled charging scenario.

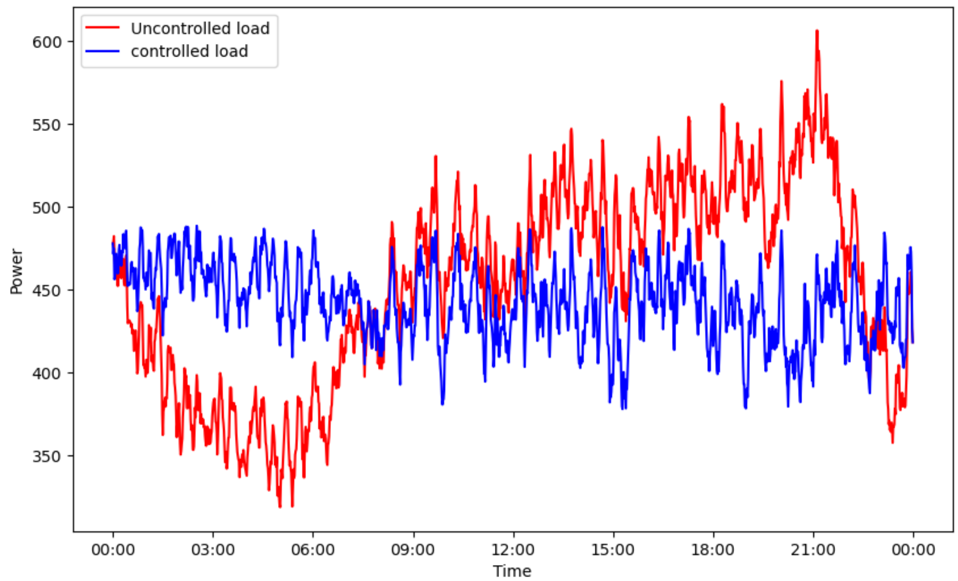

Figure 4-9: Uncontrolled load (red) and controlled load (blue) for peak load optimization only.

Finally, the net impact of both uncontrolled and controlled EV charging is significantly affected by the proportion of the total system load that the EV load comprises. For residential loads, this is directly related to the degree of EV penetration and the prevalence of home charging. Lower amounts of EV penetration can be simulated by decreasing the number of daily charging sessions or applying the same charging sessions to a larger system. Figure 4-10 shows the results of the latter approach, where the base residential load is simply scaled up by a factor of 5. While actual load aggregation is not simply multiplicative, the results offer a reasonable approximation of a larger system. In this case, the total peak load is more strongly affected by the underlying residential load, and a scheduling routine has more flexibility to charge electric vehicles during off-peak hours without creating a new peak demand window. With equal weighting of cost and peak demand minimization, the controlled schedule increases peak demand by only 1%, and the total cost of EV charging is $73. This cost is 5% smaller than the charging cost in the smaller residential system, due to the increased scheduling flexibility. Compared to the uncontrolled scenario, the controlled schedule offers a 17% reduction in EV charging cost, and a 0.3% lower peak demand. Relative to the total cost and demand of the system, these improvements over the uncontrolled charging are much smaller than for the higher EV penetration scenario.

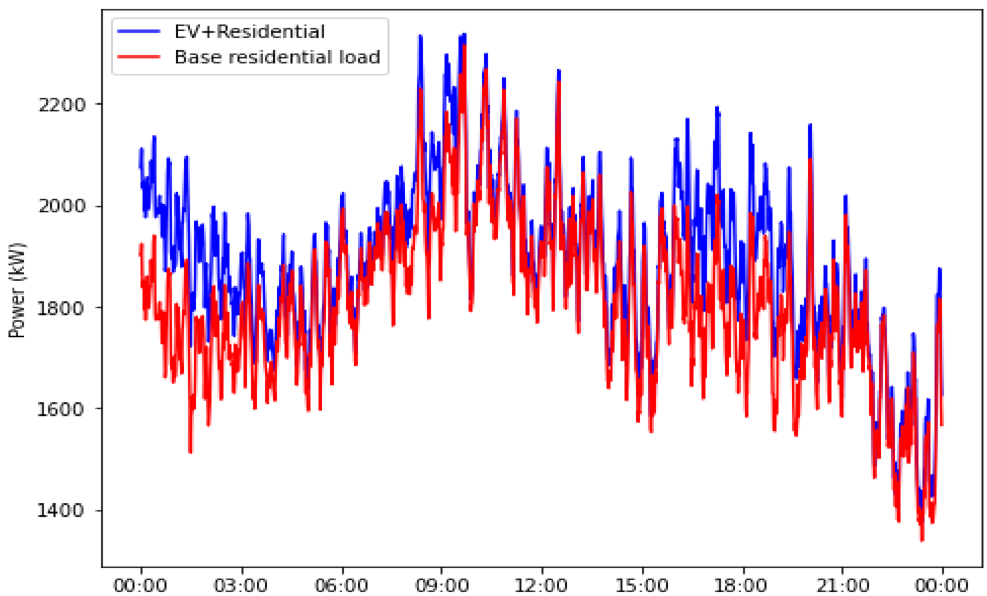

Figure 4-10: Base residential load (red) and total controlled load (blue) for a larger residential system (lower EV penetration) with equal cost and peak demand weights.

4.4 Conclusions

This chapter has yielded significant quantitative findings that shed light on its benefits and implications. One of the key takeaways is that EV charging scheduling presents a win-win solution for both electric vehicle owners and utilities. By strategically managing charging sessions, it is possible to reduce the overall cost of electricity for EV owners while minimizing the impact of charging on the electricity grid. This finding highlights the potential for coordinated scheduling to create a more efficient and sustainable charging ecosystem.

An optimization model was specifically designed for single-day scheduling of electric vehicle charging. This model was rigorously tested using real-world charging behavior data, ensuring its practical relevance. By applying the model to residential systems of varying sizes, the study examined the trade-off between minimizing electricity costs and reducing peak demand. The results demonstrate that the optimization model can strike a balance between these two objectives, providing insights into the optimal allocation of charging resources.

Furthermore, the chapter explored the scalability of controlled charging as EV penetration increases. The findings revealed an increasing performance improvement of controlled charging compared to uncontrolled charging as more electric vehicles enter the market. This suggests that as EV adoption grows, the benefits of implementing scheduling strategies become more pronounced, enabling efficient resource allocation and minimizing strain on the electricity grid.

The findings presented in this chapter lay the foundation for future investigations into the integration of prediction models for EV charging behavior in scheduling, unlocking the significant potential for optimizing the charging process. By harnessing the power of these models, scheduling algorithms can make informed decisions by considering anticipated demand, resulting in cost minimization, reduced peak loads, and enhanced overall efficiency of the charging infrastructure. The upcoming chapter will delve deeper into the realm of EV charging demand prediction, providing a comprehensive exploration of the advantages and practical implications of this approach within the context of scheduling optimization.

5. EV CHARGING DEMAND PREDICTION

5.1 Overview

In the previous chapter, the intriguing realm of EV charging scheduling was investigated, unveiling the intricate interplay between analyzed charging behavior and the creation of efficient scheduling routines. Many current charge scheduling approaches assume perfect knowledge of future prices and user charge behavior, which limits their practical applicability. In practice, this information cannot be known with certainty – however if user behavior follows consistent patterns, it could be predicted to some degree of accuracy. In order to optimize the charging schedule of electric vehicles (EVs) while meeting their charging requirements, the starting time, duration of stay, and energy requirements of each EV are crucial parameters for optimization. These parameters rely on predictions to minimize human involvement. However, the uncertainty associated with EV user behavior makes it challenging to accurately predict these parameters.

Within this chapter, attention is directed towards harnessing the power of machine learning models to delve deeper into the realm of EV user behavior prediction. By employing advanced techniques, the aim is to accurately forecast the charging behaviors of EV users. These predicted values have the potential to optimize the EV charging scheduling process, aiming for a balance between electricity demand, grid stability, and user preferences. By integrating machine learning algorithms, an opportunity arises to provide a powerful means to uncover hidden patterns, trends, and dependencies within EV charging data that may elude traditional statistical methods. These algorithms excel at capturing intricate relationships between charging parameters and user characteristics, enabling more accurate predictions and deeper insights into charging patterns.

Through these models, not only can the charging behaviors of individual EV users be anticipated, but valuable insights into aggregate charging demands at different time intervals can also be gained. The culmination of these efforts lies in the integration of EV charging demand prediction with the existing EV charging scheduling framework. By incorporating the predicted charging demands into the scheduling algorithm, the allocation of charging resources can be optimized, peak loads minimized, and the overall efficiency of the charging infrastructure enhanced.

While in this chapter, the predictive framework has been applied ambitiously to data from many different household charging stations, the same framework could be applied to data from a smaller area or even a single station, in which the input parameters may have an even higher correlation to the charging behavior, resulting in better predictions for a smaller subset of users. The feature space considered is small enough, and the algorithms fast enough, for implementation in a dynamic real-time model that continually learns from user behavior and updates future predictions.

5.2 EV Charging Behavior Prediction Framework

Data are collected and analyzed from available Level 2 charging stations located throughout Omaha, NE from Apr 2018 to Dec 2022 as shown in Table 3-1. The charging stations are single phase 40A, 240V with single charging ports. Initially, the dataset was limited to the period between January 2019 and December 2022, resulting in a total of 265,340 charging sessions obtained from 362 household charging stations, and for each session, the following information is collected: the unique EV user ID, start and end time, connection duration, charging duration, and kWh consumed.

Before delving into the specifics, it is essential to grasp the fundamental distinctions between session-level and area-level data. Session-level data pertains to individual charging sessions and their associated variables, while area-level data provides an overview of aggregate demand across specific timeframes or multiple locations. The focus here lies in studying the behavior and characteristics of individual charging sessions, aiming to predict session-level variables such as energy demand, connection duration, charging duration, and time for the next session. These predictions have significant implications for applications such as charging station management, scheduling optimization, and vehicle-to-grid integration.

By concentrating on session-level data, valuable insights can be uncovered, enabling informed decisions at the charging station level. This level of granularity allows for precise resource allocation, catering to the specific requirements of electric vehicle users and optimizing the charging process. Furthermore, by combining predictions of session behavior with information about the temporal and spatial distribution of sessions within an area, more accurate daily demand predictions can be generated. This comprehensive approach provides a deeper understanding of the charging ecosystem and support.

The primary objective of this chapter is to predict and understand the charging behavior of electric vehicle users during individual charging sessions. To achieve this, a set of carefully selected parameters is utilized as input, as detailed in Table 5-1. These parameters capture a wide range of information, including historical charging patterns and temporal factors. By incorporating these variables into the predictive models, valuable insights can be uncovered regarding the charging behavior of users and its implications for various applications.

The selection of specific output variables for prediction, as shown in Table 5-2, is guided by their significance in understanding and managing charging behavior. These variables include energy demand, connection duration, charging duration, and time for the next session. Predicting these outputs offers several advantages and opens up possibilities for optimizing charging station operations, scheduling, and integration with the electric grid.

Understanding energy demand is crucial for estimating power requirements and effectively allocating charging resources. Understanding energy demand patterns allows for the optimization of resource allocation and the efficient utilization of charging infrastructure. Prediction of connection duration and charging duration helps optimize the availability of charging stations and enables better service management for users. Anticipating the time for the next session provides insights into user behavior patterns and aids in scheduling future charging activities. Overall, accurate predictions of these variables enhance the overall charging experience, support charging infrastructure management, and facilitate the effective integration of electric vehicles into the existing grid infrastructure.

To address these research objectives, a carefully designed methodology is employed. The chosen approach involves leveraging machine learning techniques to develop robust models capable of effectively learning from the available dataset and accurately predicting the charging behavior outputs. The subsequent analysis and evaluation of these models will provide valuable insights into the factors influencing charging behavior, the relationships between input parameters and output variables, and the overall performance of the predictive models in real-world scenarios.

Table 5-1: Parameters of interest for each charging session.

Parameters	Symbol	Type	Description
Time Seq	(T_s)	Numeric	The absolute time series of the session start
Time of Day	(T_d)	Numeric	The time of day when the electric vehicle plugs in
Time Elapsed	(TE_S)	Numeric	Time elapsed since the last recorded charge ended
Cumulative Frequency	(C_f)	Numeric	The count of previous sessions for each user
Average Frequency	(A_f)	Numeric	The average daily charge frequency for each user
Session Order	(S_o)	Numeric	The sequence of a charging session for each user
Previous Value	(P_s)	Numeric	The value that precedes in a sequence for each user
User Energy Max	$(E_{S\,max})$	Numeric	The energy max for each user, for previous sessions
User Energy Mean	$(E_{S\,mean})$	Numeric	The energy mean for each user, for previous sessions
User Energy Min	$(E_{S\,min})$	Numeric	The energy min for each user, for previous sessions
User Connection duration Max	$(Co_{S\,max})$	Numeric	The connection duration max for each user, for previous sessions
User Connection Duration Mean	$(\mathrm{Co}_{S\,mean})$	Numeric	Cumulative connection duration mean for each user
User Connection duration Min	$(\mathrm{Co}_{S\,min})$	Numeric	Cumulative connection duration min for each user
User Charging duration Max	$(Ch_{S\,max})$	Numeric	Cumulative charging duration max for each user,
User Charging duration Mean	$(\mathrm{Ch}_{S\,mean})$	Numeric	Cumulative charging duration mean for each user
User Charging duration Min	$(\mathrm{Ch}_{S\,min})$	Numeric	Cumulative charging duration min for each user
Time for next session Max	$(Tn_{S\,max})$	Numeric	Cumulative time for next session max for each user
Time for next session Mean	$(\mathrm{Tn}_{S\,mean})$	Numeric	Cumulative time for next session mean for each user
Time for next session Min	$(\mathrm{Tn}_{S\,min})$	Numeric	Cumulative time for next session min for each user
Day of the week	(D_w)	Categorical	Mon, Tue, Wed, Thu, Fri, Sat, and Sun
Month of the year	(M_y)	Categorical	Jan, …. , Dec
Season	(S_y)	Categorical	Winter, Spring, Summer, and Fall

Table 5-2: Targets for each charging session.

Parameters	Symbol	Type	Description
Charging Demand	(E_s)	Numeric	The energy consumed during the charging session in kWh
Connection Duration	(Co_s)	Numeric	The connection duration of the charging session in minutes
Charging Duration	(Ch_s)	Numeric	The charging duration of the charging session in minutes
Time Until Next Charge	(Tn_s)	Numeric	The time for next charging session in minutes

The unique user ID is not used as a variable, in order to explore the dependence of charging behavior on available statistics of an arbitrary user, rather than find a functional relationship specific to each user. This approach potentially yields lower accuracy than user-specific modeling, but is much more easily generalized to large populations, fast enough for real-time prediction applications, and allows for the exploration of charging behavior patterns that are common between users. Instead, for each session, key statistics are calculated based on past user behavior, including the mean, maximum, and minimum values for energy ($E_{S\,mean}, E_{S\,max}, E_{S\,min}$), connection duration ($Co_{S\,mean}, Co_{S\,max}, Co_{S\,min}$) charging duration ($Ch_{S\,mean}, Ch_{S\,max}, Ch_{S\,min}$), and time for next session ($Tn_{S\,mean}, Tn_{S\,max}, Tn_{S\,min}$), Additional factors considered are the accumulated sum of frequencies or counts as values progress in a dataset (C_f), The average daily charge frequency for each user (A_f), the sequence of a charging session within a series of sessions for a particular user (S_o), the most recent preceding value before the current point (P_s), and the time elapsed since the last session ended (TE_s). These statistics provide valuable insights into user charging behavior and its corresponding output variables.

The prediction of charging demand $(\widehat{E_s})$ can thus be expressed as a function of these twelve parameters, shown in equation (1)

$$(\hat{E}_s) = f(\mathrm{Rid},\ Rid_s,\ T_s,\ T_d,\ \mathrm{TE}_S,\ C_S,\ D_w\ ,\ S_y\ ,\ H_y\ ,\ W_w,\ E_{smin},\ E_{smean},\ E_{max}) \qquad (1)$$

The prediction of connection duration $(\widehat{Co_s})$ can thus be expressed as a function of these twelve parameters, shown in equation (2)

$$(\widehat{Co_s}) = f(Rid,\ Rid_s,\ T_s,\ T_d,\ TE_S,\ C_S,\ D_w,\ S_y,\ H_y,\ W_w,\ Co_{smin},\ Co_{smean},\ Co_{max}) \qquad (2)$$

The prediction of charging duration $(\widehat{Ch_s})$ can thus be expressed as a function of these twelve parameters, shown in equation (3)

$$(\widehat{Ch_s}\ ) = f(Rid,\ Rid_s,\ T_s,\ T_d,\ TE_S,\ C_S,\ D_w\ ,\ S_y\ ,\ H_y,\ W_w,\ Ch_{smin},\ Ch_{smean},\ Ch_{max}) \qquad (3)$$

The prediction of time for the next session $(\widehat{Tn_s})$ can thus be expressed as a function of these twelve parameters, shown in equation (4)

$$(\widehat{Tn_s}) = f(Rid,\ Rid_s,\ T_s,\ T_d,\ TE_S,\ C_S,\ D_w,\ S_y,\ H_y\ ,\ W_w,\ Tn_{smin},\ Tn_{smean},\ Tn_{max}) \qquad (4)$$

The objective of this chapter is to assess the feasibility of predicting the charging behavior, using only information available at the start of charging session. If the output is assumed to be a function of the input parameters in Equation 1, 2, 3, and 4 the inputs and outputs of this function are known for every session in the dataset. Regression analysis can then be used to approximate an underlying function that maps a given set of input parameters (the information known at charging) to the output parameter (the recorded energy demand, connection duration, charging duration, and time for the next session). This approximated function (model) can then be used to predict the output of future sessions, based on the input parameters of those sessions. The overall framework is illustrated in Figure 5-1.

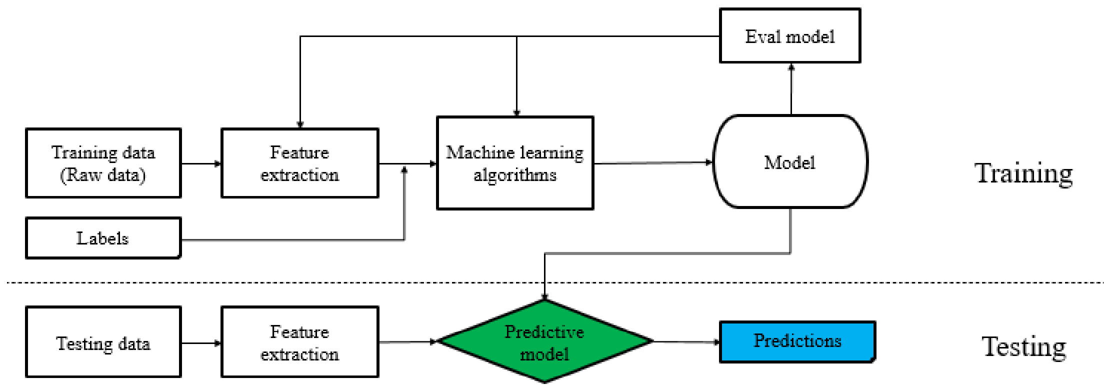

Figure 5-1: Charging demand prediction framework.

Three machine learning algorithms, namely Gradient Boosting (XGBoost), Random Forest (RF), and Artificial Neural Network (ANN), are investigated for predicting charging behavior. These algorithms strike a balance between accuracy and computational speed, making them suitable for real-time applications. A detailed explanation of these methods is provided in the following subsection.

Machine learning models face several challenges that need to be addressed. One such challenge is the risk of overfitting, where the model becomes overly complex and fits the training data too closely, leading to poor generalization of new, unseen data. To ensure reliable predictions, it is crucial to carefully tune and validate the machine learning models, mitigating the risk of overfitting. Another consideration is the interpretability of machine learning model results. While these models can provide accurate predictions, their inner workings can be complex and difficult to interpret compared to traditional statistical methods. This complexity can make it challenging to gain insights into the specific relationships and interactions between variables in the model.

In the context of this chapter, linear regression is used as a reference point. However, linear regression assumes a linear relationship between the dependent and independent variables, which may not capture the explicitly nonlinear relationships expected in many variables, especially categorical ones. Consequently, the linear regression model may have limitations in capturing the full complexity of the data and may yield less accurate predictions compared to more advanced machine learning methods.

5.3 Machine Learning Methods

Supervised machine learning is a type of machine learning in which the algorithm learns to predict outputs based on labeled input data. In supervised learning, the dataset is divided into a set of inputs (also known as features) and outputs (also known as labels). The algorithm is trained on this labeled dataset to learn a mapping function from the input variables to the output variables. Once the algorithm is trained, it can then be used to predict the output for new, unseen input data.

The goal of supervised learning is to minimize the difference between the predicted output and the actual output for the training data. The algorithm uses a cost function to measure the difference between the predicted output and the actual output, and then adjusts the model parameters to minimize this cost. There are many established regression algorithms, with various advantages and disadvantages such as :

5.3.1 Gradient Boosting

Boosting frameworks are often chosen due to their effortlessness and extraordinary outcomes on average-size datasets. XGBoost, in particular, has seen widespread use in data science due to its high accuracy, flexibility, speed, and efficiency [103]. It is used to solve regression, classification, and ranking problems [104]. XGBoost's concept is to improve the performance of computational power for boosted tree algorithms. This algorithm is considered to be one of the fastest to incorporate tree ensemble approaches, using information from all data points in a leaf to decrease the search space of potential feature splits [35], [105].

5.3.2 Random Forest

Random forests, also known as random decision forests, are a highly utilized ensemble training method. It is commonly applied for both classification and regression and functions by building an aggregation of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees' leverage [106]. Ensemble methods use multiple learning models to gain better predictive results. In the case of a random forest, the model creates an entire forest of random uncorrelated decision trees to arrive at the best possible answer. Random forest aims to overcome the correlation issue by picking only a subsample of the feature space at each split. Fundamentally, it aims to de-correlate the trees and cut the trees by setting stopping criteria for node splits. Random forest algorithm offers an excellent accuracy among current algorithms, and runs efficiently on large datasets. It can manipulate thousands of input variables without variable deletion. It creates an inner straight estimate of the generalization error as the forest building progresses [107].

5.3.3 Artificial Neural Network (ANN)

Artificial Neural Network (ANN), also known as a neural network or simply a neural net, is a type of computational model that is inspired by the structure and function of the human brain. ANN is a form of deep learning, which is a subset of machine learning, and it is widely used for a variety of tasks, including image and speech recognition, natural language processing, and data analysis[108].

An ANN is composed of interconnected nodes or neurons organized into layers. The basic unit of an ANN is a neuron, which receives inputs, applies an activation function, and produces an output [109]. Neurons are organized into layers, with the input layer receiving the input data and the output layer producing the final output or prediction [110]. In between the input and output layers, there can be one or more hidden layers, which help to capture complex patterns in the data.

During training, an ANN learns to make predictions by adjusting the weights associated with the connections between neurons. This is done through a process called backpropagation, where the error between the predicted output and the actual output is used to update the weights [111]. The ANN continues to iterate through this process until the error between the predicted and actual outputs is minimized.

ANNs are capable of learning complex patterns from large amounts of data and can generalize well to make predictions on unseen data. They are capable of handling both linear and non-linear relationships in data, and their architecture can be customized to suit the specific problem at hand. However, ANNs can be computationally expensive and require a large amount of data for training, and they may also suffer from overfitting if not properly regularized. Nevertheless, ANNs have become a popular and powerful tool in various fields of machine learning and artificial intelligence.

5.4 Machine Learning Methods' Accuracy Evaluations

A model's accuracy is evaluated by examining the differences between the predictions of the model and the actual observations in the test set. Because there are thousands of observations in the test set, these differences are summarized by common statistical evaluation metrics, and these metrics are compared for each of the four regression methods. The following subsection explains more about the evaluation metrics used in this research:

5.4.1 Coefficient of Determination (R^2)

R^2 is an important performance metric for any regression analysis. Used in statistical models for many applications, it provides a quantification of how well the model predicts the relationship between the input data and the generated output. A model that always generates a perfect prediction would have an R^2 of one, while a model whose predictions do not respond at all to input parameters would have an R^2 of zero.

The coefficient of determination, R^2, can be mathematically defined using equation (2). In this equation, the numerator represents the sum of squares of the residuals (SS_{RES}), while the denominator corresponds to the sum of squares for the test set (SS_{TOT}). Interpreting R2 as a ratio of variances provides insight into the proportion of variance in the result that is explained by the input parameters.

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} = \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y}_i)^2} \qquad (2)$$

Where, y_i is the actual value from the test set, $\hat{y}_i$ is the predicted value of y_i and $\bar{y}_i$ is the mean of the y_i values.

5.4.2 Root Mean Square Error (RMSE)

Root Mean Square Error (RMSE) is another common statistical metric, quantifying the average amount of error between a prediction and a test set. RMSE has the same units as the variable being predicted. It is defined by equation (3) and is simply the standard deviation of the residuals or errors. RMSE provides information on how far, on average, a model's predictions are from their expected values.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \qquad (3)$$

Where n is the number of observations.

5.4.3 Mean Absolute Error (MAE)

Like the RMSE, mean absolute error (MAE) is also commonly used to quantify the average amount of error between a prediction and a test set. Instead of calculating the standard deviation of residuals, the MAE is simply the average of the absolute value of the residuals, as seen in equation (4). While RMSE and MAE are similar, RMSE gives a higher weight to larger errors before averaging. When the MAE is significantly lower than the RMSE, it can indicate a larger spread in the values of the residuals.

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n} \qquad (4)$$

5.5 Data Processing and Splitting

Before feeding the machine learning model, it is essential to perform data cleaning and data splitting to ensure the quality and reliability of the analysis. Data cleaning encompasses a series of crucial procedures aimed at addressing missing values, outliers, and inconsistencies within the dataset. When encountering missing values, an appropriate course of action involves either imputing them with estimated values or removing them altogether, depending on the degree of missingness and the characteristics of the variable in question. Outliers, which are extreme values that deviate significantly from the rest of the data, may need to be identified and either treated or removed if they are deemed influential or erroneous. Inconsistencies in the data, such as conflicting or illogical entries, should be resolved through data validation and verification procedures.

In this chapter, a series of data cleaning procedures were conducted to ensure the integrity and reliability of the dataset consisting of 265,340 sessions. Specific criteria were applied to select sessions for inclusion while excluding those that did not meet the defined requirements. These procedures were implemented to enhance the quality of the dataset and ensure that the subsequent analysis and modeling processes are based on reliable and relevant data. Consequently, sessions that consumed 1 kilowatt-hour (kWh) or less were excluded from the analysis as they typically indicated connection errors or technical issues with the charging station. Similarly, sessions with a connection duration of 5 minutes or less were also removed for the same reason. After applying these criteria, the dataset was reduced to 241.040 remaining valid charging sessions.

In order to improve the accuracy and reliability of the analysis, a subset of the original dataset was selected by excluding sessions that were deemed as outliers or potential errors. The goal was to focus on a more representative and reliable subset of users for modeling purposes.

To achieve this, specific criteria were applied to filter the sessions. Sessions from users with durations exceeding 7 days or with gaps of more than 7 days between sessions were eliminated. This was done to exclude extreme outliers that may not align with typical charging patterns or could be indicative of data anomalies.

Furthermore, to ensure an adequate sample size for meaningful analysis, sessions from users with at least 500 sessions or more were included in the final dataset. This decision was made to prioritize users with a substantial charging history, providing a more comprehensive basis for analysis while mitigating potential biases associated with users with limited data.

After applying the selection criteria, the final dataset consisted of 157,374 sessions. These sessions were from users who met the established criteria for duration and frequency of charging sessions, ensuring a more focused and reliable dataset for further analysis.

In addition to the data cleaning procedures mentioned earlier, a restriction was placed on the outputs of the dataset to further refine the analysis. Specifically, the outputs were limited to the 95th percentile of the total values. This restriction was implemented to mitigate the influence of outliers and extreme values that may skew the analysis and affect the interpretability of the results. Figure 5-2 depicts the scatterplot of the original output values before applying the restriction to the 95th percentile. This scatterplot provides an overview of the entire range of output values, allowing for visual inspection of the distribution and potential presence of outliers.

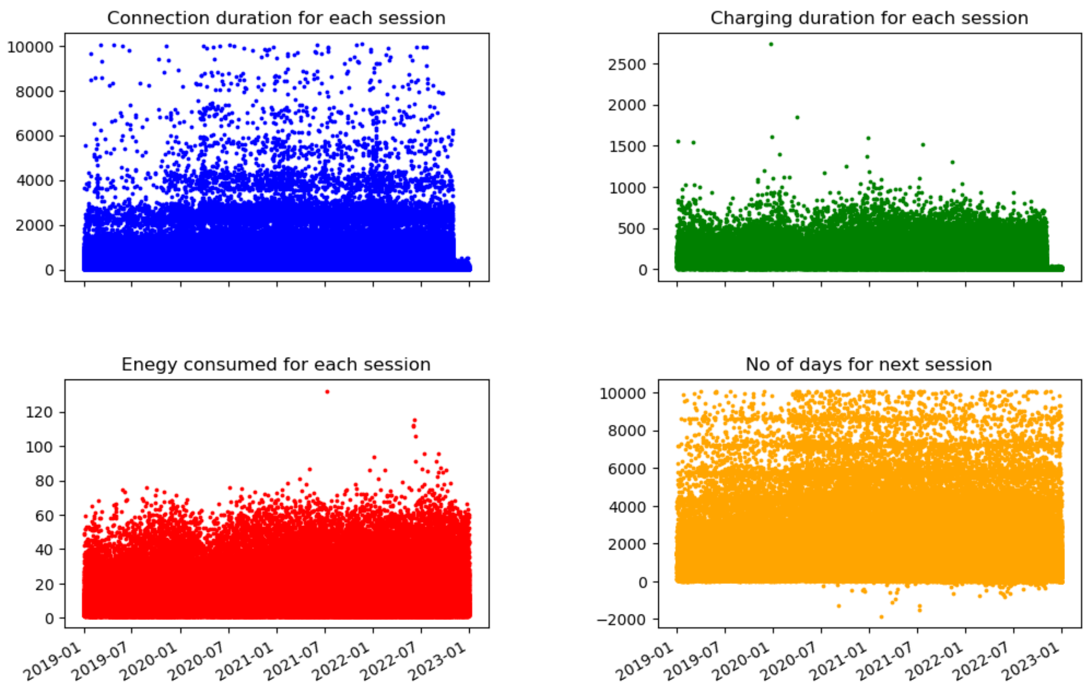

Figure 5-2: Scatterplot of the original output values before applying the restriction to the 95th percentile.

Similarly, Figure 5-3 shows the corresponding histogram of the output values after removing the outliers. This histogram provides a more detailed representation of the distribution, highlighting the concentration of values within certain ranges and the overall shape of the data distribution.

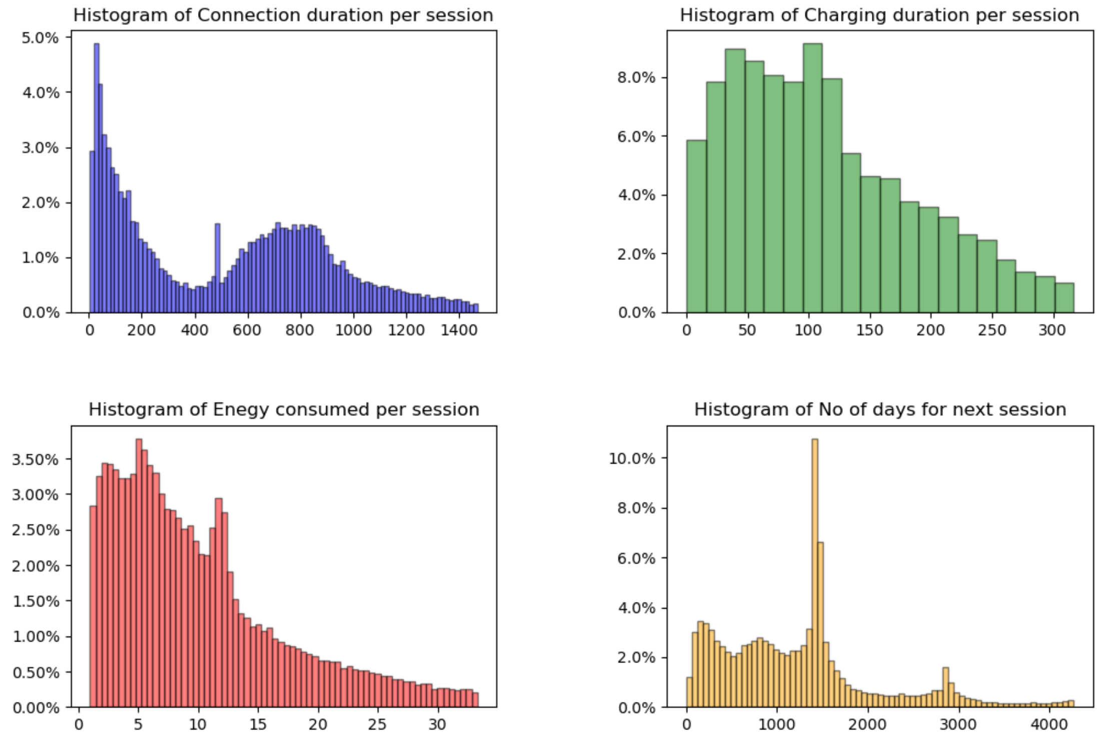

Figure 5-3: Histogram of the output values after removing the outliers.

Table 5-3 presents a summary of statistical measures for the target variables, providing valuable insights into their distribution and variability. The measures include the minimum value, first quartile (25th percentile), mean, third quartile (75th percentile), maximum value, and standard deviation.

The minimum value represents the smallest observed value of the target variable, indicating the lower bound of its range. The first quartile represents the value below which 25% of the data points fall, providing information about the lower end of the distribution. The mean represents the average value of the target variable, giving a measure of its central tendency. The third quartile represents the value below which 75% of the data points fall, providing information about the upper end of the distribution. The maximum value represents the largest observed value of the target variable, indicating the upper bound of its range.

The standard deviation measures the dispersion or variability of the target variable around the mean. It indicates the spread of the data points and how much they deviate from the average value. A higher standard deviation suggests a greater degree of variability, while a lower standard deviation indicates less variability.

For instance, in "Connection Duration," the minimum value is 5, indicating the shortest duration observed in the dataset. The median value of 531 indicates that 50% of the sessions in the dataset had a connection duration less than or equal to this value, while the remaining 50% had a longer duration. It serves as a central point that divides the distribution of connection durations into two halves. The mean value of 509 represents the average connection duration across all sessions. The 1st and 3rd quartiles (123 and 814, respectively) give us a sense of the spread of the data, with half of sessions falling within this range. The maximum value of 1468 indicates the longest connection duration observed. The standard deviation of 387 reflects the variability in connection durations, with higher values indicating a wider dispersion of data points around the mean.

The variables related to EV charging sessions exhibit different distribution patterns. Energy and charging duration variables follow fairly standard one-sided distributions, while connection duration and time until the next session have distinct distributions. Additionally, all variables have standard deviations that are a high percentage of their means, indicating a significant relative spread in values. These characteristics present a challenge for accurately predicting and optimizing these variables in the context of EV charging schedules.

Table 5-3: Summary of the distribution of the charging behavior targets.

	Min	**1st Qu**	**Median**	**Mean**	**3rd Qu**	**Max**	**Std**
Connection Duration	5	123	531	509	814	1468	387
Charging Duration	0	52	100	122	160	316	74
Energy	1	4.6	8.3	10	13.4	33.2	7.2
Time next charge	2	585	1200	1232	1496	4264	855

To address the issue of overfitting, a careful division of the dataset into training and testing sets was employed. This division strategy serves to mitigate the risk of overfitting and assess the model's capacity to generalize to unseen data.

The training set, encompassing 80% of the data, is dedicated to training the model. During this phase, the model acquires an understanding of the underlying patterns and relationships in the data. By optimizing its parameters, the model minimizes errors and enhances its performance.

Following the model's training, its performance is assessed using the testing set, which comprises the remaining 20% of the data. This testing set consists of new, unseen data that the model has not been exposed to during training. By evaluating the model's performance on this independent dataset, a reliable estimation of its generalization capabilities is obtained. A strong performance on the testing set indicates the model's effectiveness in making accurate predictions on novel, unseen data.

The test/train split strategy plays a crucial role in mitigating overfitting by subjecting the model to a rigorous evaluation with unseen data. This approach enables the validation of the model's ability to generalize beyond the training data, ensuring that the conclusions drawn from its performance are dependable and applicable in real-world scenarios.

The allocation of sessions to each subset is a critical decision, as it determines the data on which the model is trained and tested. A time-based split, for instance, would enable the model to learn from past charging behavior and make predictions for future behavior. However, it is essential to consider scenarios where the dataset contains multiple users with different charging patterns over time.

In order to address this concern and facilitate a comprehensive learning process for the model while also testing its performance against "future" behavior, the following steps are implemented.

1. The dataset is sorted by user, and the initial session of each user is discarded. This discarded session serves as the starting point for calculating key parameters such as the mean, maximum, and minimum energy demand of previous sessions, as well as the time elapsed since the last charge.
2. The first 80% of charging sessions for each user, based on chronological order, are assigned to the training set. This ensures that the model learns from a substantial portion of each user's charging history.
3. The subsequent 20% of charging sessions for each user are allocated to the testing set. This segment is used to evaluate the model's performance on unseen data, mimicking "future" charging behavior.

It is important to acknowledge that the evaluation and performance of the model in this study are specifically focused on users with an extensive charging history. The testing set, comprising 20% of the charging sessions for each user, serves as a representation of "future" charging behavior that the model has not been exposed to during training. It should be noted that the model's performance on this testing set reflects its convergence towards predicting the behavior of users with similar characteristics and charging patterns as those included in the dataset. Therefore, the results obtained from this evaluation are indicative of the model's potential performance in a real-world application where it is continuously updated based on user charging behavior. However, it is important to consider the limitations of this setup. The model's ability to robustly predict charging behavior for users without an extensive charging history or those with significantly different patterns may not be fully evaluated in this study. The focus is on mimicking a specific real-world scenario where the model is continually trained and updated based on user data. Therefore, while the model's performance is assessed and evaluated based on the convergence observed during training and testing, it is crucial to interpret the results within the context of the specific application and user population for which the model is intended.

The implementation of each model in this study is done using the Python programming language which offers a comprehensive set of libraries and tools for machine learning tasks. Python's popularity in data analysis and modeling is due to its rich ecosystem. The code for model implementation is organized and executed within Jupyter Notebook, an interactive environment that facilitates code development, visualization, and documentation. For model training and evaluation, several packages from the Scikit-learn library are utilized [112]. Scikit-learn provides convenient functions for training models and tuning parameters for Linear, XGBoost [113], Random Forest [114], and Artificial Neural Network (ANN) [115] methods.

5.6 EV Charging Behavior Prediction Results

The Results section analyzes and interprets the performance of the implemented models in predicting the charging behavior outputs. The models were trained using the training dataset and tested using the test dataset. This section presents the evaluation metrics and statistical measures employed to assess the predictive performance of each model.

The accuracy metrics, such as mean absolute error (MAE), root mean squared error (RMSE), and R-squared, are reported. These metrics offer insights into the overall predictive accuracy and goodness of fit of the models. Lower values of MAE and RMSE indicate better accuracy, while a higher R-squared value suggests a better fit of the model to the data.

Graphical representations of the model predictions, such as scatter plots, are provided to compare the predicted values with the actual target values. These visualizations offer a visual assessment of how well the models capture the underlying patterns and trends in the data. The residual plots are included in the results to detect any systematic patterns or biases in the model's predictions. Ideally, the residuals should exhibit a random scatter around zero, indicating an unbiased prediction with an equal balance of overestimation and underestimation. However, deviations from this pattern may indicate specific ranges or patterns of the target variable where the model performs better or worse.

The analysis of feature importance plots in the results section allows for the identification of influential variables that have a significant impact on the target variables. These plots provide valuable insights into the relationships and dynamics within the dataset, revealing which variables play a crucial role in the model's predictive performance. Typically, feature importance is assessed using metrics such as the Gini index or information gain, which quantify the contribution of each variable. The presentation of the results is done through bar charts or ranked lists, emphasizing the variables with the highest importance.

In the results section, Recursive Feature Elimination (RFE) is utilized. The purpose of employing RFE in the study is to improve the feature selection process. RFE is an iterative method that aims to identify the most relevant features by repeatedly eliminating less important ones based on their rankings or importance scores. By performing RFE, researchers can gain insights into the relative importance of different features in predicting the target variable. The iterative nature of RFE allows for a systematic exploration of feature subsets and helps identify the optimal set of features that contribute the most to the predictive accuracy of the model. Furthermore, RFE provides a way to control the complexity of the model by selecting a desired number of features. This trade-off between the number of features and model performance can be visualized and analyzed, enabling researchers to make informed decisions about feature selection.

5.6.1 Charging Demand (Energy)

Accuracy metrics for predicting charging demand in kilowatt-hours (kWh) are presented in Table 5-4, showcasing the performance of different methods. RF achieved the highest R^2 value of 48%, indicating a better fit to the data compared to other methods. XGBoost and ANN also performed with R^2 values of 46% and 40%, respectively, while Linear regression had the lowest R^2 value of 29%. However, it should be noted that RF's performance, although the highest, still does not provide accurate predictions as it only accounts for roughly 48% of the variability in the charging demand.

Considering MAE and RMSE, which measure the average and overall prediction errors, respectively, RF exhibited the lowest values among all methods, indicating better accuracy in predicting charging demand. XGBoost and ANN also showed relatively low MAE and RMSE values, suggesting good performance in estimating charging demand. Linear regression, on the other hand, had higher MAE and RMSE values compared to the other methods, indicating less accurate predictions. However, it is important to note that the overall MAE and RMSE values across all methods are still relatively high, indicating room for improvement in achieving more accurate charging demand predictions.

When comparing the results of the machine learning models with the mean of each user, it becomes apparent that the models exhibit comparable performance to basic statistical analysis. In this approach, the mean value for each user is calculated and then compared to the actual value of each session.

Through the evaluation of machine learning models, the effectiveness of predicting charging behaviors can be assessed by comparing their performance to the mean charging behavior of each user. For instance, the mean of each user yields an R-squared value of 38% and an RMSE (Root Mean Squared Error) of 5.7 kWh.

By contrasting these baseline results with the performance of the machine learning models, it becomes evident that the models demonstrate comparable or superior performance. The accuracy metrics achieved by the models, such as higher R-squared values and lower RMSE values, indicate their ability to capture and predict charging behaviors more accurately than relying solely on the mean values of each user.

The STD and Mean values in the table are additional measures of the prediction errors' variability and bias, respectively. These values are consistent across all methods, indicating that the average prediction errors and their dispersion are similar for all models.

Table 5-4: Accuracy Metrics for predicting charging demand in kWh.

Methods	R^2	MAE	RMSE	STD	Mean
Linear	0.31	4.7	6	7.2	10.2
XGBoost	0.46	3.9	5.3	7.2	10.2
RF	0.48	3.6	5.1	7.2	10.2
ANN	0.40	4.1	5.5	7.2	10.2
Mean	0.38	4.3	5.7	7.2	10.2

Figure 5-4 showcases a visual representation of the alignment between the predicted and actual values, enabling an assessment of their accuracy. The plot serves as evidence that the random forest (RF) method outperforms other methods in predicting charging demand, as indicated by the relatively close proximity between the predicted and actual values on the plot.

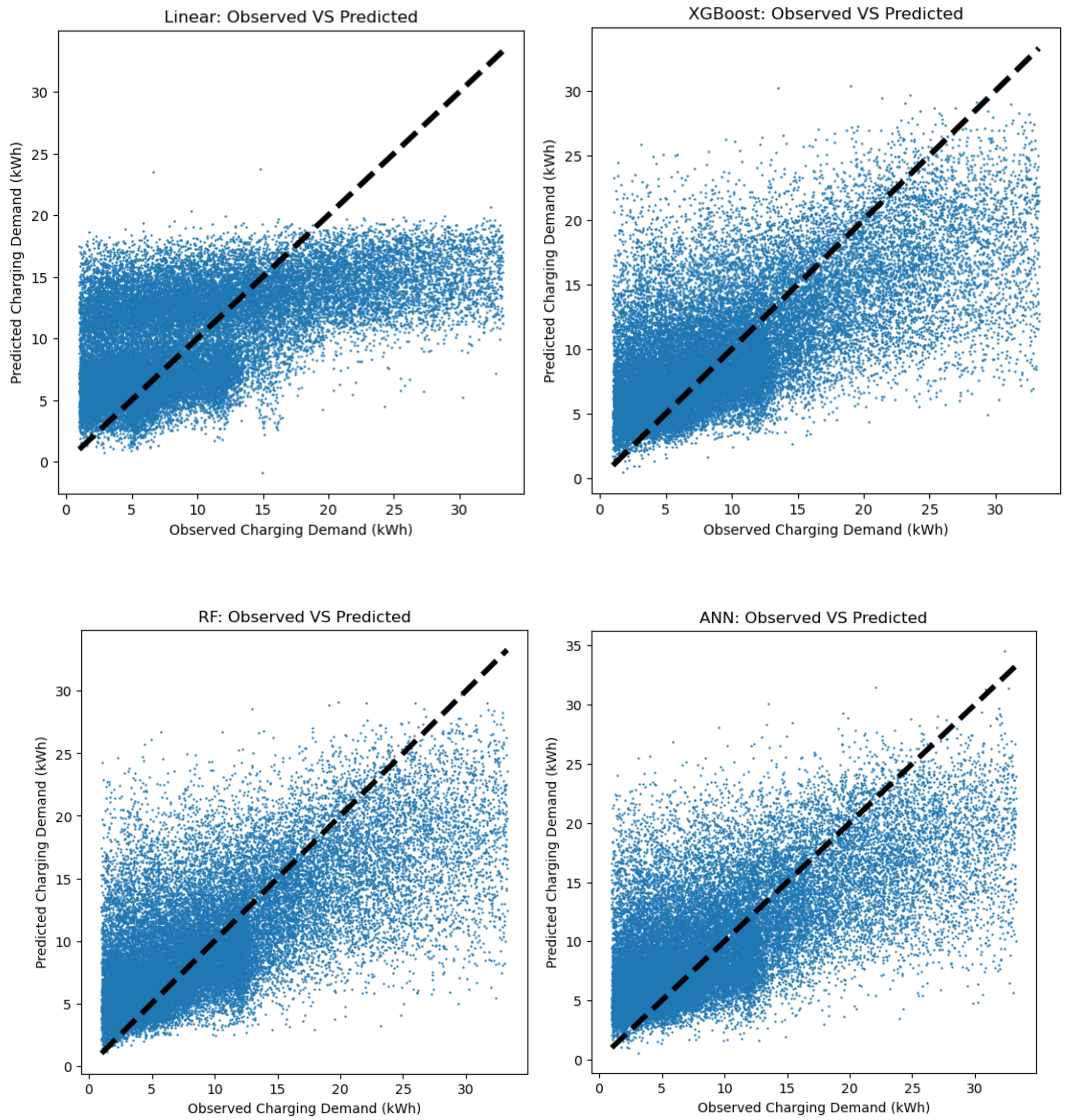

Figure 5-4: Predicted vs. Observed charging demand for each method.

Furthermore, Figure 5-5 displays the residual plot for each method, revealing discernible patterns and unique characteristics. Across all the methods, including Random Forest (RF), the residuals appear to be randomly scattered around zero. This indicates that the models achieve a balanced representation of overestimation and underestimation, suggesting their ability to capture the underlying patterns in the data. Consequently, these methods demonstrate relatively accurate predictions of energy consumption compared to other approaches.

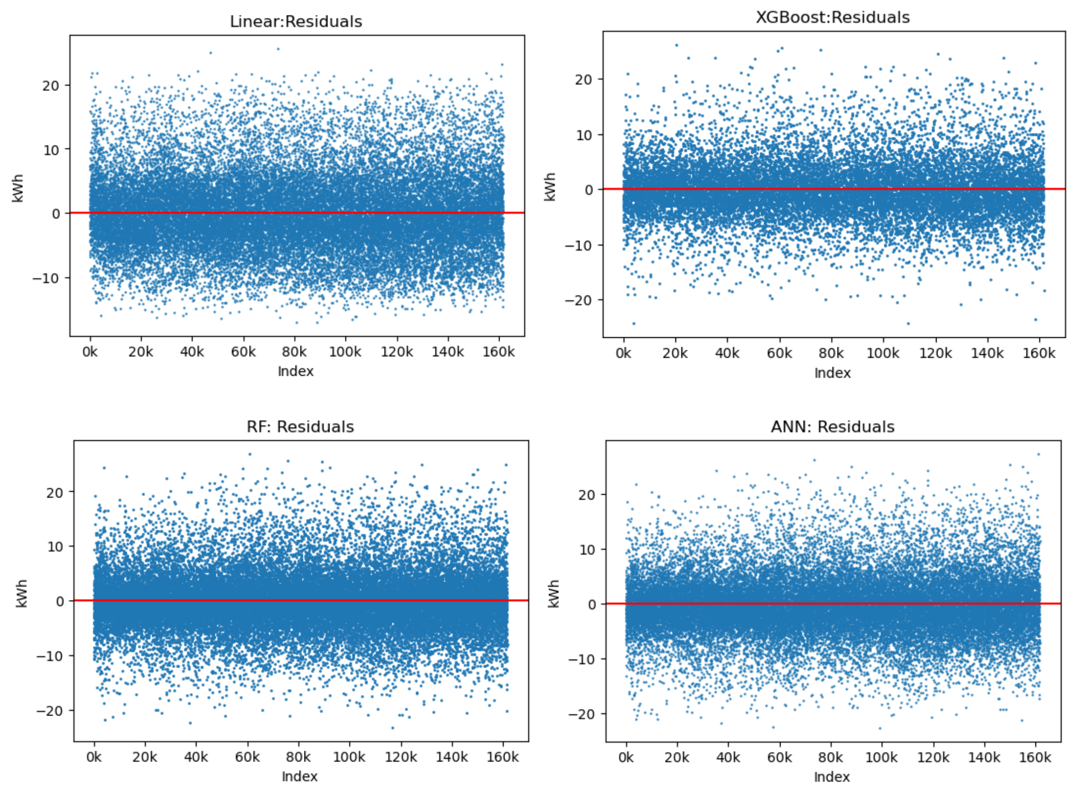

Figure 5-5: Residuals of charging demand predictions for each method.

To gain a deeper understanding of the factors influencing the prediction of energy consumption concerning the different variables used for classifying charging sessions, a feature importance analysis is carried out. Figure 5-6 presents the feature importance plots for each method, shedding light on the variables that have the most significant impact on energy consumption predictions. In the case of the Random Forest (RF) method, the feature importance plot indicates that the mean is the most important variable in determining energy consumption, followed by other significant variables such as time of the day, time elapsed, and others. These variables play a crucial role in the prediction of energy consumption and contribute significantly to the overall accuracy of the RF model.

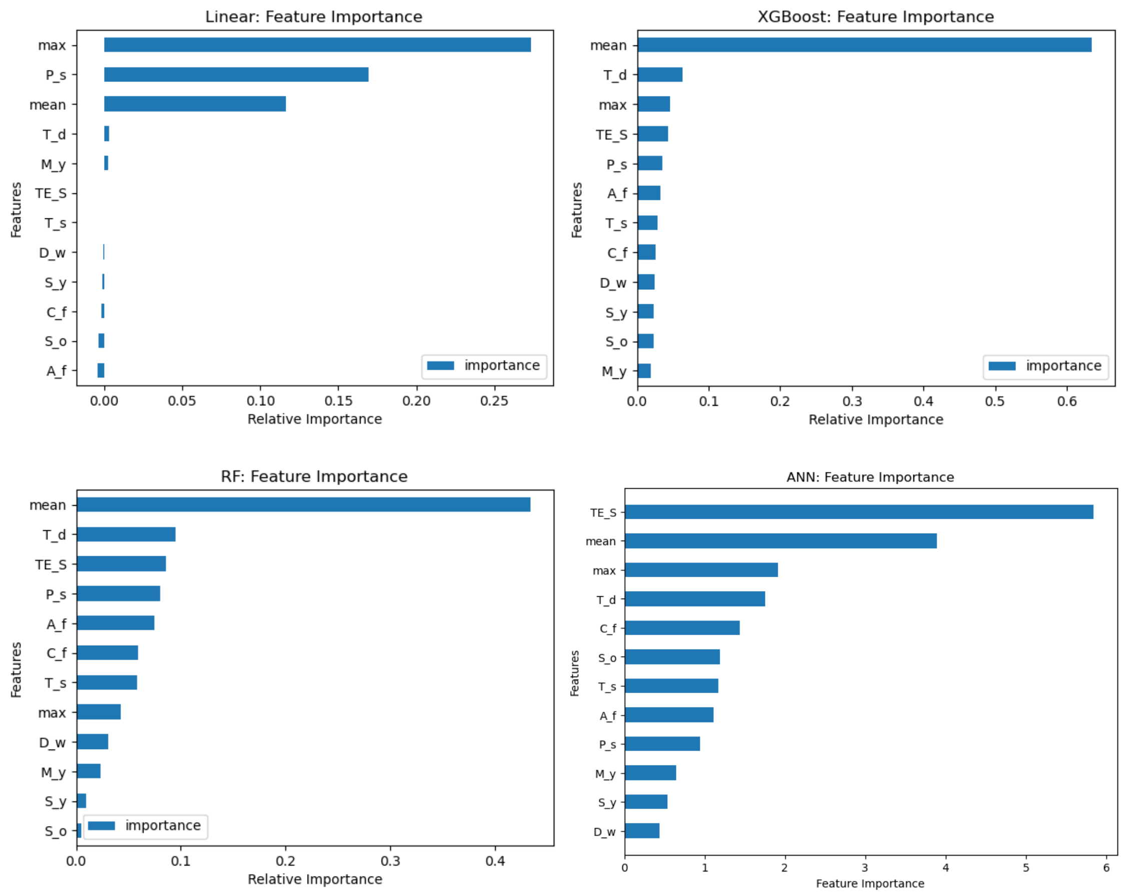

Figure 5-6: Feature Importance of charging demand predictions for each method.

In Figure 5-7, the RFECV plot reveals the significance of different features in predicting charging demand with a cross-validation fold of 5. The plot demonstrates that the first 4-6 features have the highest importance, as indicated by the maximum adjusted r^2 score. Beyond these features, the plot shows diminishing improvements in the model's performance with the addition of more variables.

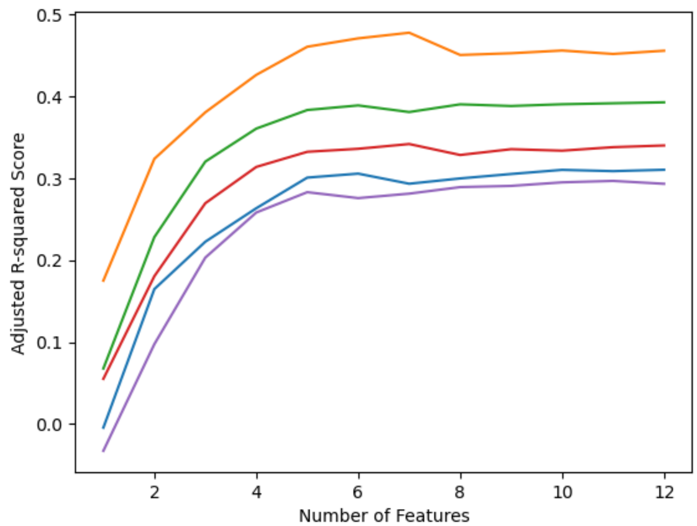

Figure 5-7: Optimal feature selection for predicting charging demand using RF method: RFECV Analysis.

Additionally, Table 5-5 highlights the importance of specific variables, including mean, average frequency, time of day, time elapsed, previous charging demand value, and absolute time series. These six features consistently contribute to a higher adjusted R-squared value, indicating their strong influence on the accuracy of the predictive model for charging demand. The ranking indicates the order of importance of each feature in forecasting the charging demand.

Table 5-5: Feature Importance of charging demand predictions using RF method.

Rank	**Features**
1	Mean
2	Average Frequency
3	Time of Day
4	Time Elapsed
5	Previous Charging Demand Value
6	Absolute Time Series

5.6.2 Connection Duration

Among the evaluated machine learning methods used to predict the connection duration, Random Forest (RF) emerged as the most promising and reliable approach. The RF algorithm's ability to harness the power of decision trees proved beneficial in capturing both linear and non-linear relationships between the input features and the connection duration. The Random Forest (RF) algorithm outperformed other machine learning methods in predicting connection duration, as evidenced by lower values of mean absolute error (MAE) of 209 minutes and root mean square error (RMSE) of 301 minutes. These metrics indicate that RF provided more accurate and precise estimates of the connection duration, despite the modest R^2 value of 41% as shown in Table 5-6.

Table 5-6: Accuracy Metrics for predicting connection duration.

Methods	R^2	MAE	RMSE	STD	Mean
Linear	0.20	282	346	386	511
XGBoost	0.38	218	307	386	511
RF	0.40	210	305	386	511
ANN	0.34	227	313	386	511
Mean	0.12	304	362	386	511

Figure 4-8 presents the comparison of predicted and observed connection duration for various machine learning methods, aiming to evaluate their performance. Notably, the random forest (RF) method exhibits superior predictive accuracy in connection duration compared to other methods.

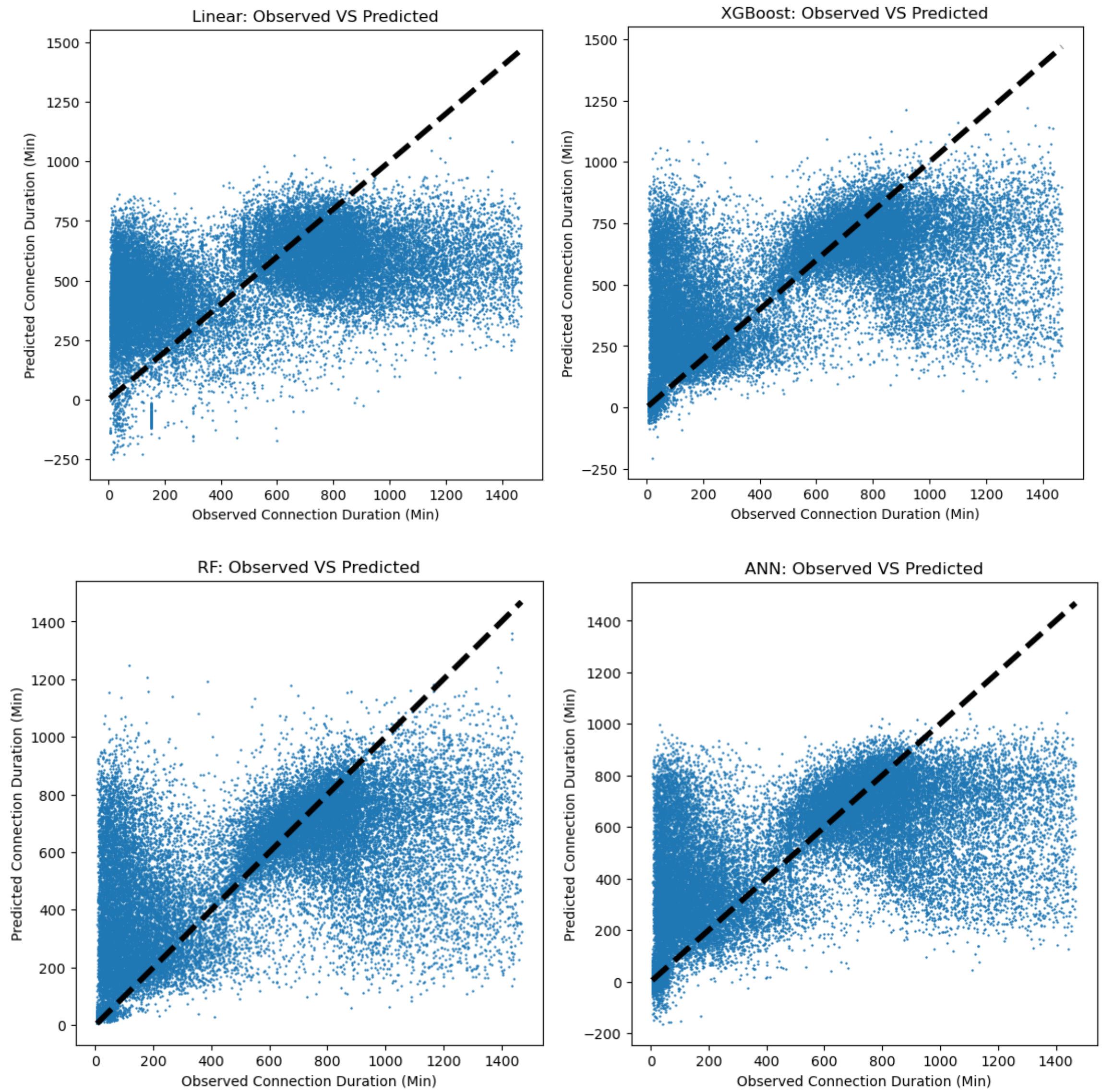

Figure 5-8: Predicted vs. Observed connection duration for each method.

Figure 5-9 illustrates the residual plot for each method, revealing unique patterns and characteristics. Notably, in the case of the Random Forest (RF) method, the residuals exhibit a random distribution around zero. This suggests that the RF model achieves a better-balanced representation of both overestimation and underestimation, indicating its ability to capture underlying patterns in the data. Consequently, the RF method demonstrates a higher level of accuracy in predicting connection duration compared to other methods.

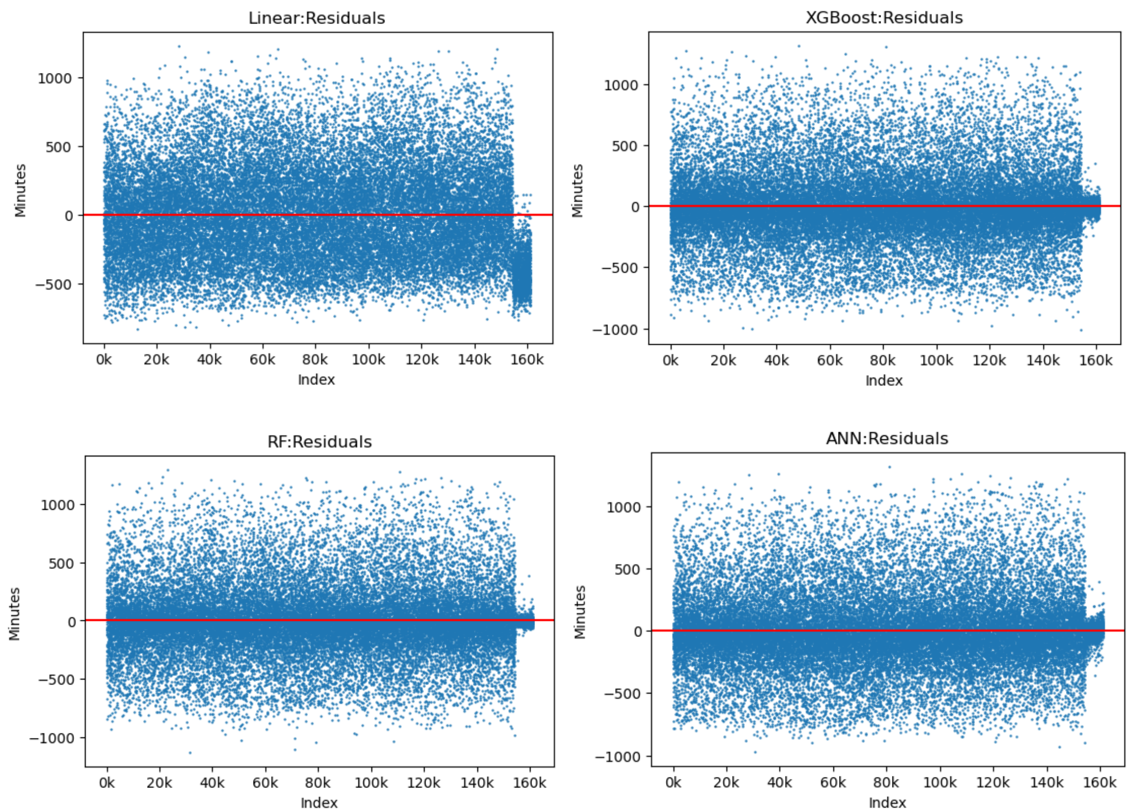

Figure 5-9: Residuals of connection duration for each method.

In order to explore the factors influencing the prediction of connection duration concerning the various variables used for classifying charging sessions, a feature importance analysis is conducted. Figure 5-10 showcases the feature importance plots for each method, providing insights into the variables that play a significant role in connection duration predictions. Specifically, for the Random Forest (RF) method, the feature importance plot reveals the specific variables that carry substantial importance in determining connection duration. These variables may encompass mean, time of the day, time elapsed, and other relevant factors.

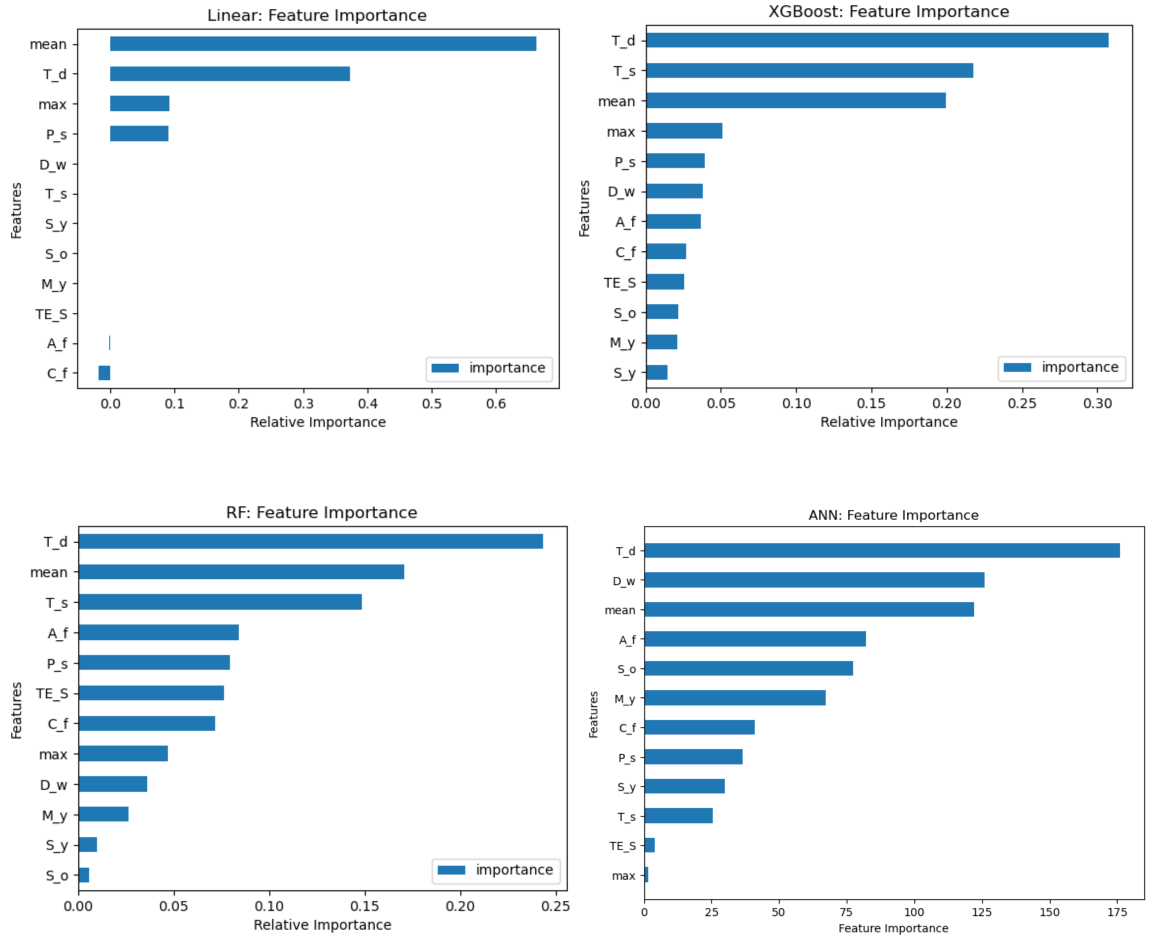

Figure 5-10: Feature Importance of connection duration prediction for each method.

The RFECV plot in Figure 5-11 illustrates the importance of various features in predicting connection duration using a cross-validation fold of 5. It is evident from the plot that the initial 4-6 features exhibit the highest significance, as indicated by the maximum adjusted r^2 score. However, beyond these features, the plot demonstrates diminishing improvements in the model's performance with the inclusion of additional variables.

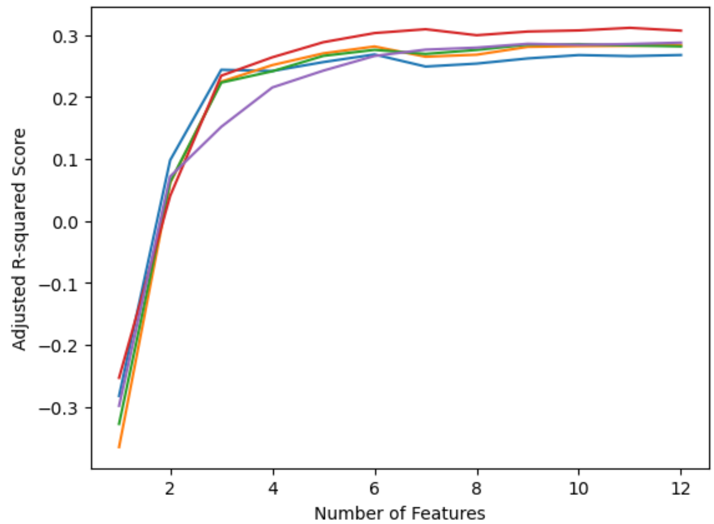

Figure 5-11: Optimal feature selection for predicting connection duration using RF method: RFECV analysis.

Table 5-7 presents the significance of certain variables, such as mean, time of day, absolute time series, average frequency, previous connection duration value, and time elapsed, in accurately predicting the connection duration. These six features consistently demonstrate a strong influence on the accuracy of the predictive model for connection duration, as reflected by their higher adjusted R-squared values. The ranking of these variables provides valuable insights into their relative importance in effectively forecasting the connection duration.

Table 5-7: Feature Importance of connection duration predictions using RF method.

Rank	**Features**
1	Mean
2	Time of Day
3	Absolute Time Series
4	Average Frequency
5	Previous Connection Duration Value
6	Time Elapsed

5.6.3 Charging Duration

Table 5-8 summarizes the evaluation metrics for each method, including R^2, MAE, and RMSE, to assess their performance in predicting the duration of EV charging sessions. The Random Forest (RF) model outperforms other methods with an R^2 value of 47% and the lowest MAE and RMSE values of 40 and 54, respectively. The XGBoost method also shows promising performance, while the Linear and ANN methods exhibit lower R^2 values and higher MAE and RMSE values. This suggests that the XGBoost model is capable of capturing the complexity of the data and generating reliable predictions of charging duration. In contrast, the Linear and ANN methods exhibit relatively lower R^2 values and higher MAE and RMSE values, indicating potential limitations in capturing the intricate relationships within the data. However, the superior performance of the RF and XGBoost models underscores the effectiveness of advanced machine learning techniques in predicting charging duration for EV charging scheduling.

Table 5-8: Accuracy Metrics for predicting charging duration.

Methods	R^2	MAE	RMSE	STD	Mean
Linear	0.29	48	62	74	112
XGBoost	0.45	41	55	74	112
RF	0.47	40	54	74	112
ANN	0.40	43	57	74	112
Mean	0.26	50	64	74	112

Figure 5-12 provides a comparison between the predicted and observed charging duration for different machine learning methods, to evaluate their performance. The random forest (RF) method demonstrates greater predictive accuracy in charging duration compared to the other methods. This conclusion is supported by the proximity between the predicted and actual values depicted in the plot.

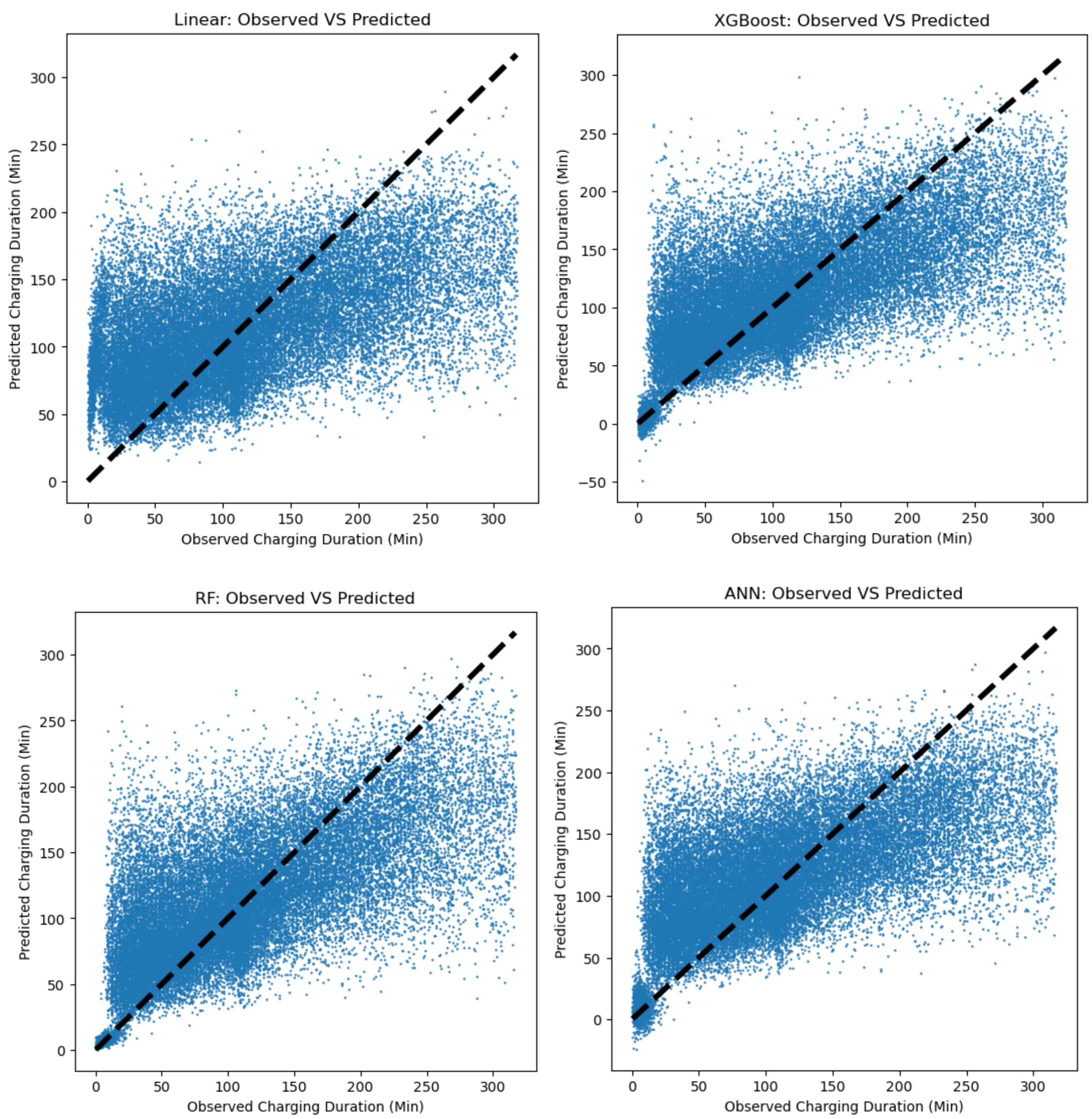

Figure 5-12: Predicted vs. Observed charging duration for each method.

Additionally, the residuals plot depicted in Figure 5-13 showcases the distinct patterns and characteristics observed for each method. Particularly, for the Random Forest (RF) method, the residuals display a random distribution centered around zero. This characteristic indicates that the RF model effectively captures the underlying patterns in the data, resulting in a more balanced representation of both overestimation and underestimation. As a result, the RF method outperforms other methods in terms of accuracy when predicting charging duration.

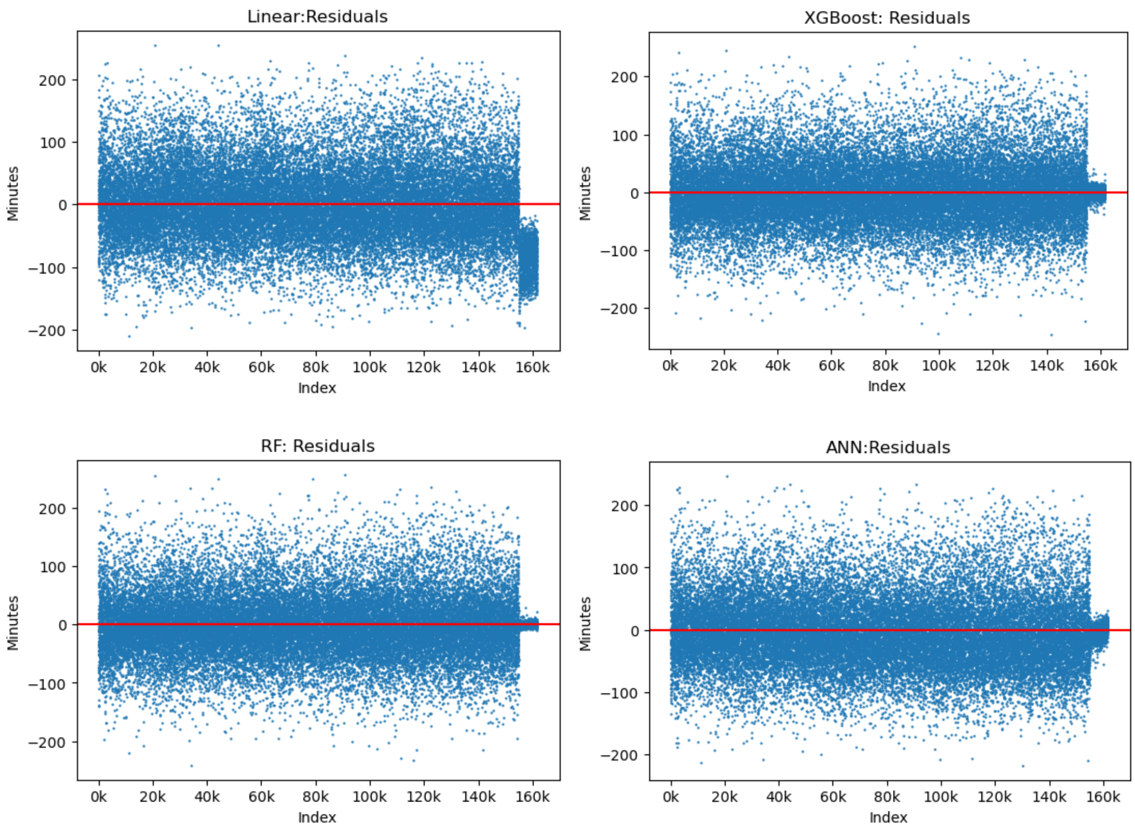

Figure 5-13: Residuals of charging duration for each method.

To investigate the factors that impact the prediction of charging duration and their relationship with the variables used for classifying charging sessions, a feature importance analysis was performed. The feature importance plots for each method are presented in

Figure 5-14, shedding light on the variables that hold significant importance in predicting charging duration. Notably, for the Random Forest (RF) method, the feature importance plot highlights specific variables such as mean, time of the day, time elapsed, and other relevant factors that greatly influence connection duration.

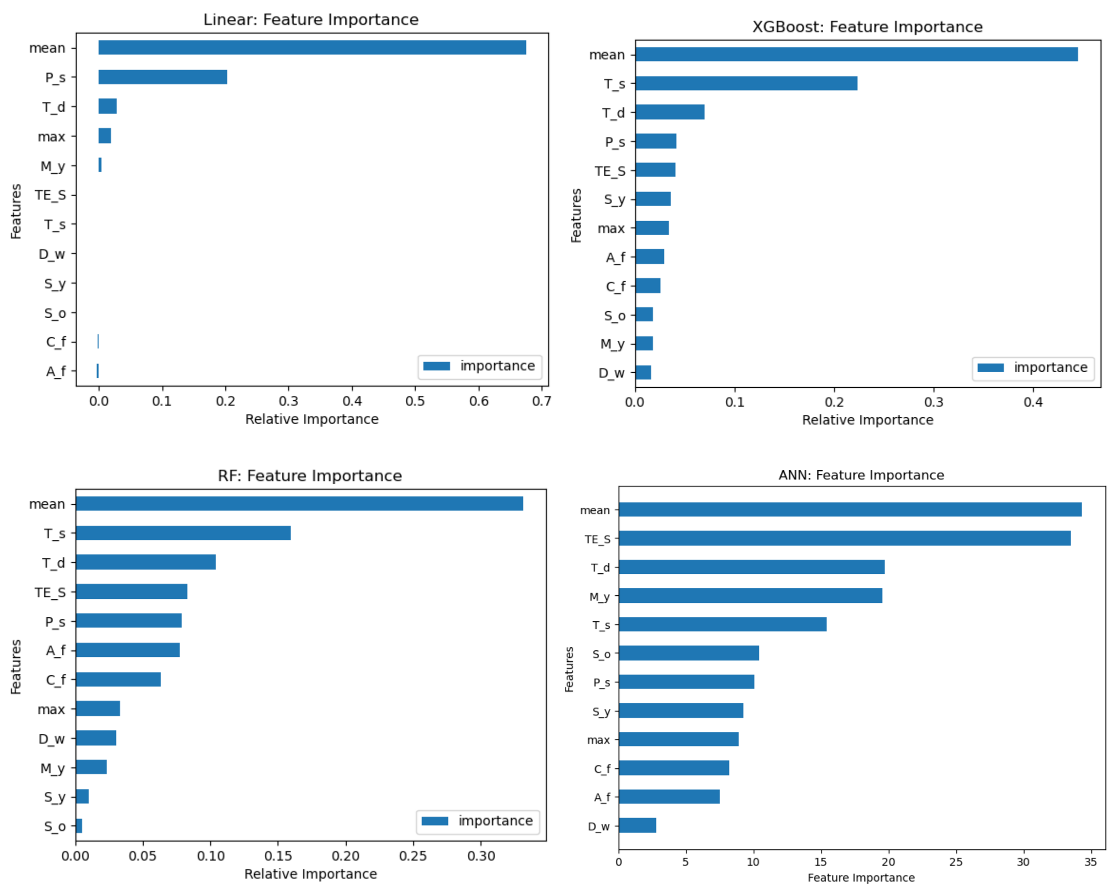

Figure 5-14: Feature Importance of charging duration predictions for each method.

Figure 5-15, showcasing the RFECV plot, provides insights into the significance of different features in predicting charging duration with a cross-validation fold of 5. The plot highlights that the first 4-6 features hold the highest importance, as denoted by the maximum adjusted r^2 score.

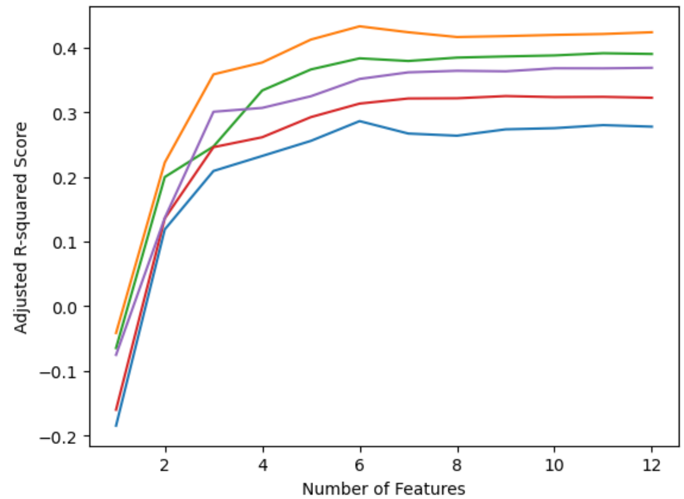

Figure 5-15: Optimal feature selection for predicting charging duration using Random Forest: RFECV analysis.

Table 5-9 provides an overview of the significance of specific variables, including mean, absolute time series, time of day, average frequency, time elapsed, and previous charging duration value, in accurately predicting the charging duration. These six features consistently exhibit a strong influence on the accuracy of the predictive model, as evident from their higher adjusted R-squared values. The ranking of these variables offers valuable insights into their relative importance in effectively forecasting the charging duration.

Table 5-9: Feature Importance of charging duration predictions using RF method.

Rank	Features
1	Mean
2	Absolute Time Series
3	Time of Day
4	Average Frequency
5	Time Elapsed
6	Previous Connection Duration Value

5.6.4 Time Until Next Charge

In the specific task of predicting the time until the next charge, the XGBoost algorithm exhibited relatively better performance compared to other machine learning methods. However, it is important to note that even with this algorithm, the results were still unsatisfactory with R^2 equaling to only 21%. Additionally, the mean absolute error (MAE) and root mean square error (RMSE) values were relatively high as shown in Table *5-10*: Accuracy Metrics for predicting time for the next session in min. These results indicate that the model's predictions deviated significantly from the actual values, suggesting limitations in capturing the underlying patterns and dynamics of the time for the next charge.

Table 5-10: Accuracy Metrics for predicting time for the next session in min.

Methods	**R^2**	**MAE**	**RMSE**	**STD**	**Mean**
Linear	0.16	583	781	864	1234
XGBoost	0.21	565	760	864	1234
RF	0.20	567	761	864	1234
ANN	0.17	582	774	864	1234
Mean	0.16	585	783	864	1234

The performance of the models in predicting the time for the next charge can be visually represented in Figure 5-16, demonstrating their accuracy in forecasting. The alignment between the predicted values and the observed values serves as an indicator of the models' effectiveness. Ideally, a close correspondence between the predicted and observed values would indicate accurate predictions.

However, in this case, the plot shows a noticeable deviation between the predicted and observed values. This suggests that the models and linear regression approach might not capture the underlying patterns and relationships in the data effectively.

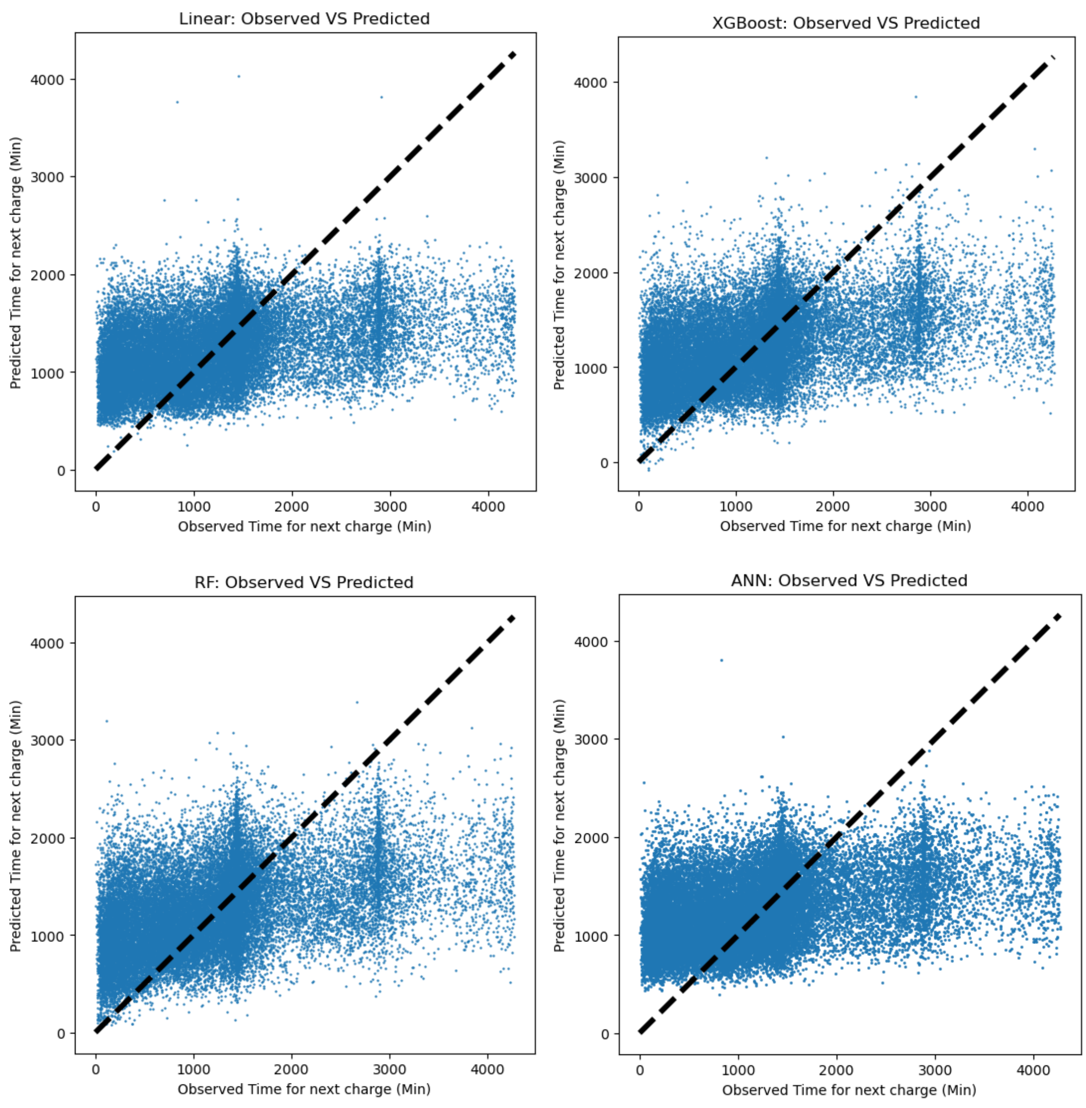

Figure 5-16: Predicted vs. Observed time for the next session for each method.

Figure 5-17 displays the residuals against the index and provides insights into the presence of any systematic patterns or biases in the models' predictions. Ideally, the residuals should be randomly scattered around zero, indicating a balanced representation of overestimation and underestimation. However, certain patterns and trends can be observed in this plot, indicating that the models and linear regression approach may not adequately capture the underlying complexity of the Time for the next charge. These patterns can provide insights into the models' performance, highlighting areas where they perform well or poorly in predicting certain ranges or patterns of the target variable.

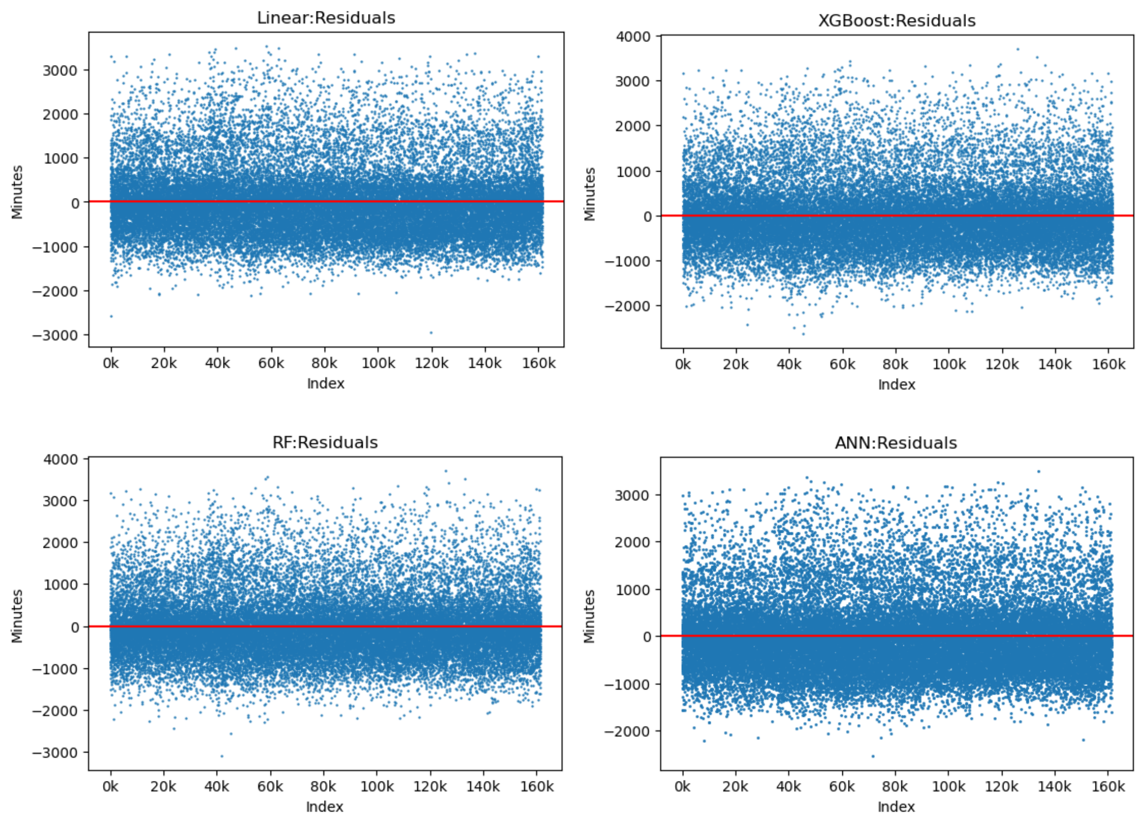

Figure 5-17: Residuals of time for the next session for each method.

To explore the factors influencing the prediction of time for the next charge and their relationship with the variables used for classifying charging sessions, a feature importance analysis was conducted. Figure 5-18 showcases the feature importance plots for each method, providing insights into the variables that play a significant role in predicting Time for the next charge. Specifically, for the Random Forest (RF) method, the feature importance plot reveals the specific variables, including mean, time of the day, time elapsed, and other relevant factors, that hold substantial importance in determining time for the next charge.

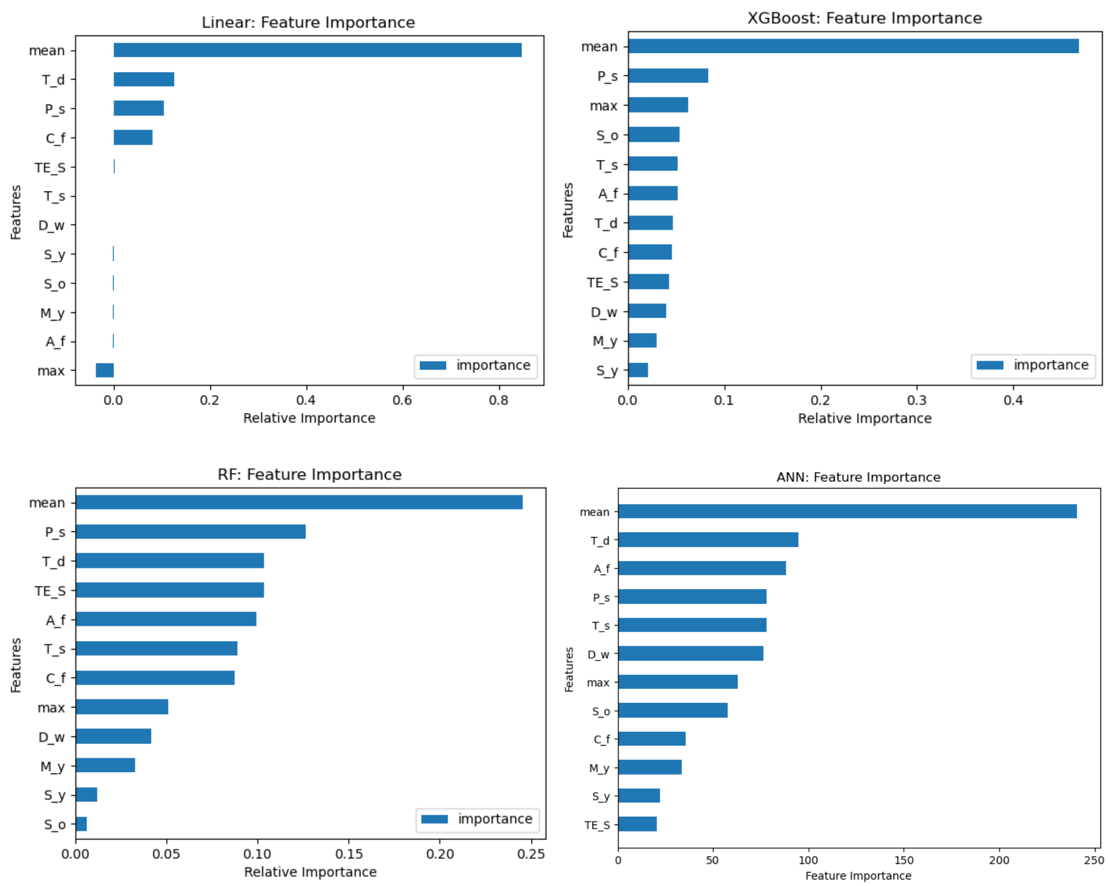

Figure 5-18: Feature Importance of time for the next session predictions for each method.

Furthermore, Figure 5-19, which displays the RFECV plot, offers valuable insights into the importance of various features in predicting the time until the next charge. By utilizing a cross-validation fold of 5, the plot indicates that the initial 4-6 features exhibit the highest significance, as evidenced by the maximum adjusted R^2 score.

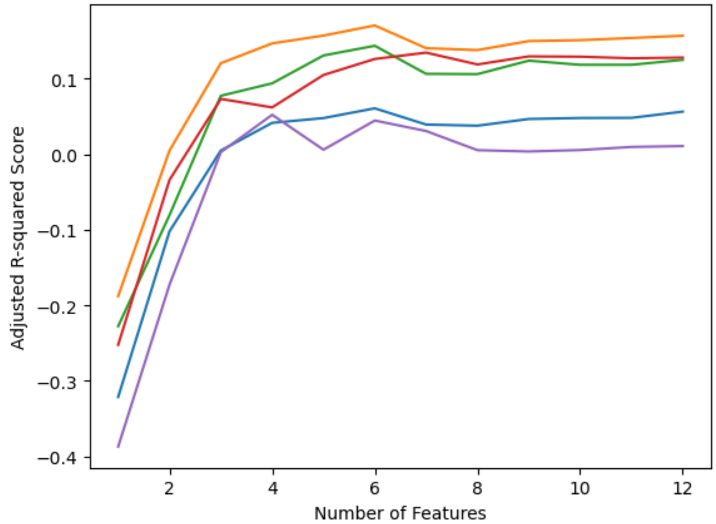

Figure 5-19: Optimal feature selection of time for next session predictions using RF method: RFECV analysis.

Table 5-11 summarizes the importance of specific variables, including mean, average frequency, previous value of time for the next charge, time elapsed, absolute time series, and time of day, in predicting the time for the next charge. These six features consistently demonstrate a strong influence on the accuracy of the predictive model, as reflected by their higher adjusted R-squared values. The ranking of these variables provides valuable insights into their relative importance in effectively forecasting the time for the next charge.

Table 5-11: Feature Importance of next session predictions using RF method.

Rank	Features
1	Mean
2	Average Frequency
3	Previous Value of time for next charge
4	Time Elapsed
5	Absolute Time Series
6	Time of Day

5.7 Conclusions

The main objective of this chapter was to investigate and compare the predictive performance of various machine learning models concerning different aspects of EV charging behavior. The models considered for analysis included Linear, Xgboost, Random Forest, and Artificial Neural Network ANN, with a focus on predicting charging demand, connection duration, charging duration, and time for the next charge. It is worth noting that these models were applied at the individual charging session level to allow for a more detailed examination of the data.

The results of predicting charging demand indicate that the Random Forest (RF) method performed the best among the evaluated models, with an R^2 value of 48%. However, it is important to note that even the best-performing model can only explain 48% of the variability in charging demand, highlighting the inherent complexity and variability in accurately predicting such behaviors. Evaluating the models in relation to the mean of each user revealed additional insights. The mean-based approach achieved an R^2 value of 38% and an RMSE of 5.7 kWh, underscoring the significance of the mean as a key factor in energy consumption. The feature importance analysis further confirmed the importance of variables such as time of the day and time elapsed. The RFECV plot provided valuable insights into the significance of different features in predicting charging demand, with mean, average frequency, time of the day, time elapsed, previous charging demand value, and absolute time series identified as crucial factors in accurate forecasting.

In the analysis of predicting connection duration, the Random Forest (RF) algorithm demonstrated superior performance compared to other models, yielding lower mean absolute error (MAE) and root mean square error (RMSE) values. Although the achieved R^2 value of 41% indicates room for improvement, the RF algorithm's ability to capture both linear and non-linear relationships contributes to more precise estimates of connection duration. The feature importance analysis highlighted the key variables, including mean, time of the day, time elapsed, and others, that significantly influence connection duration predictions.

The evaluation of different methods for predicting charging duration in EV sessions indicates that the Random Forest (RF) model outperforms other models with a higher R^2 value of 47% and lower MAE and RMSE values. Although the RF model demonstrates the best performance among the evaluated methods, it is important to note that accurately predicting charging duration remains a challenging task, as indicated by the modest R^2 value. However, the results highlight the relative superiority of the RF model compared to other methods in capturing patterns and providing more accurate predictions. The RFECV plot identifies important variables, including mean, absolute time series, time of day, average frequency, time elapsed, and previous charging duration, in accurately forecasting the charging duration. Incorporating additional variables beyond these does not significantly improve the model's performance.

The XGBoost algorithm showed relatively better performance compared to other machine learning methods in predicting the time until the next charge. However, it is important to note that the results were still unsatisfactory. The R^2 value of 21% indicates that the model's predictions deviated significantly from the actual values. The MAE and RMSE values were also relatively high, indicating limitations in capturing the underlying patterns and dynamics of the time until the next charge. The feature importance analysis, particularly for the Random Forest (RF) method, identified variables such as mean, time of the day, time elapsed, and others as significant contributors to predicting the time until the next charge. These insights can provide valuable information for further refining the prediction models and improving their accuracy.

The relatively low R-squared values and high root mean square error (RMSE) indicated the difficulty of capturing the complex dynamics of charging behavior accurately. These findings suggest the need for further research and refinement to enhance the models' predictive accuracy. Table 5-12 shows the accuracy metrics comparison for predicting the four outputs using different machine learning methods.

Table 5-12: Accuracy Metrics comparison for predicting the four outputs using different ML methods.

Outputs	**Linear**		**XGBOOST**		**ANN**		**RF**	
	R^2	**RMSE**	**R^2**	**RMSE**	**R^2**	**RMSE**	**R^2**	**RMSE**
Energy	0.32	6	0.46	5.3	0.40	5.5	0.48	5.1
Connection Duration	0.20	346	0.38	307	0.34	313	0.40	305
Charging Duration	0.29	62	0.45	55	0.40	57	0.47	54
Time Until Next Charge	0.16	781	0.21	760	0.17	774	0.20	761

6. ANALYSIS OF AGGREGATED EV USER BEHAVIOR IN A SPECIFIC AREA

6.1 Overview

Expanding on the insights gained from the previous chapter's focus on EV charging behavior prediction at the session level, this chapter focuses on analyzing the aggregated user behavior in a specific area to gain valuable insights into peak demand periods and the number of simultaneously connected vehicles. By studying EV user behavior at a geographic scale, important patterns and trends can be identified, providing crucial information for infrastructure planning, resource allocation, and policy-making.

Understanding peak demand periods is essential for effectively managing charging infrastructure and mitigating potential grid stress. By examining the distribution of charging events across different timeframes, such as day, week, or year, the analysis reveals the specific periods when EV charging demand is at its highest.

Additionally, analyzing the number of simultaneously connected vehicles provides insights into the utilization of charging infrastructure. By studying the patterns of how many vehicles are connected to charging stations at a given time, it becomes possible to assess the capacity requirements and identify potential congestion points.

In summary, the analysis of aggregated user behavior in an area yields valuable insights regarding peak demand periods and the number of simultaneously connected vehicles. These insights contribute to effective infrastructure planning, load management strategies, and the overall optimization of charging infrastructure, ensuring a smooth and reliable charging experience for EV users while minimizing grid stress and maximizing resource utilization.

6.2 Descriptive Analysis of EV User Behavior

This section focuses on the descriptive analysis of aggregated EV user behavior. To prepare the dataset, certain data-cleaning procedures were implemented. Sessions with a connection duration of less than 5 minutes were removed as they were likely caused by technical issues during the charging process. Similarly, sessions with a duration exceeding 1469 minutes were excluded as they did not represent typical EV user behavior.

Furthermore, to ensure the dataset's quality and reliability, charging durations exceeding 600 minutes were excluded. This decision was made based on the understanding that the majority of EVs typically require between 4 to 10 hours for a complete charge. Additionally, users who had less than 10 recorded charging sessions were removed from the dataset. This step was taken to eliminate users with insufficient historical data, which could potentially affect the analysis. Approximately 8.5% of the total dataset was removed as part of these data-cleaning procedures. As a result, the remaining dataset comprises 25,6488 sessions, providing a more robust foundation for conducting the subsequent analysis.

Once the data was cleaned, the focus shifted to selecting users who charged their EVs during the year 2021 and had at least one monthly charging session. This subset of users comprised 222 individuals. The analysis aimed to gain insights into their charging patterns and understand their behavior on a broader scale. Two primary aspects were examined: the number of connected EVs and the power consumed during the charging period across various time resolutions.

6.2.1 Aggregated EV User Behavior Analysis

6.2.1.1 Introduction

The purpose of this section is to investigate the aggregated charging behavior of EV users. The analysis begins by examining the number of connected EVs per minute over a year. However, to gain deeper insights and uncover potential patterns, the data is further analyzed at monthly, weekly, and daily resolutions. By exploring the charging behavior at these different time intervals, valuable insights can be obtained regarding any variations or trends in the connection behavior of EV users. This comprehensive analysis aims to provide a more nuanced understanding of EV charging patterns and identify noteworthy observations.

Moving forward, the focus will be on examining the power consumption during the charging period of EVs. The power consumed by each EV will be estimated by multiplying the binary charging duration (indicating whether a charging session occurred or not) by the charger capacity of 7.5 kW. This calculation will provide an approximation of the power consumed during each EV's charging session. Furthermore, a plot will be presented to visualize the power consumption per minute throughout the year. Additionally, the analysis will delve into monthly, weekly, daily, and hourly intervals to identify any notable trends or fluctuations in power consumption patterns. By exploring power consumption at different resolutions, valuable insights can be gained into the charging behavior of EVs and any significant findings can be uncovered.

6.2.1.2 Results

A. No. of Simultaneously Connected EVs

Figure 6-1 displays the aggregate number of connected Electric Vehicles (EVs) per minute throughout the year 2021 for 222 users. The data reveals a remarkable level of stability in EV connectivity over this period. On average, there were around 52.5 EVs connected per minute, with a standard deviation of approximately 24.3. The figure provides an overall snapshot of the EV network's activity, indicating that EVs remained consistently connected to charging stations, contributing to a reliable and efficient charging ecosystem.

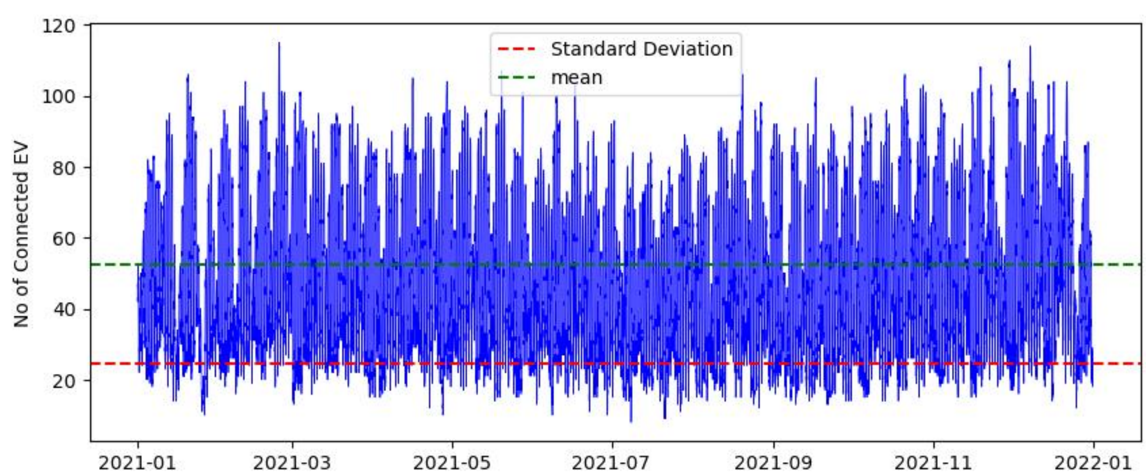

Figure 6-1: Aggregate No. of connected EV per minute.

Upon zooming out and examining the data at a monthly resolution, the aggregate number of connected Electric Vehicles (EVs) is presented alongside the corresponding standard deviation for each month in Figure 6-2. A consistent pattern emerges, showing that more EVs are connected during the winter months compared to the summer months. However, January 2021 presents an interesting exception to this trend. The lower number of connected EVs during the first month of the year can be attributed to multiple factors. One possibility is the holiday season, when fewer EVs might have been in use or connected to charging stations due to reduced travel or charging activities. Another factor could be the restriction in cleaning the data, which resulted in the exclusion of long connected sessions, particularly during the holiday period. Additionally, it is plausible that more EVs joined the network towards the end of January, given that it represents the first month of the available data. Nevertheless, the overall trend of higher EV connectivity during winter months suggests a potential relationship between weather conditions and EV usage patterns, with individuals opting for EVs more frequently during colder seasons. In terms of the standard deviation plot, it closely resembles the main plot, with minor variations. The standard deviation exhibits variation across different months, with a higher value of 25 in February and a lower value of 21.3 in July, resulting in a difference of 3.7 between the two.

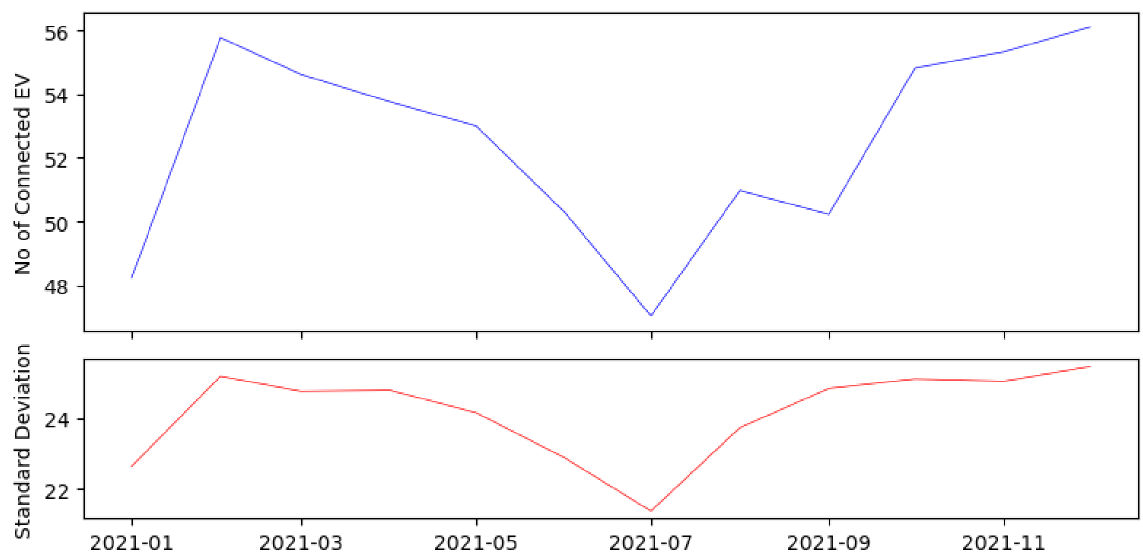

Figure 6-2: Aggregate No. of connected EV along with standard deviation each month.

Analyzing the data at a weekly resolution in Figure 6-3 reveals interesting insights into the aggregate number of connected Electric Vehicles (EVs) throughout the year 2021. Notably, the first week of the year stands out with significantly lower EV connectivity compared to the other winter weeks. This observation aligns with the previously mentioned anomaly in January 2021, suggesting a possible correlation between the two. The lower EV connectivity during the first week could be attributed to various factors, such as reduced travel or charging activities associated with the New Year's holiday. Additionally, the last week of the year exhibits a similar pattern of lower EV connectivity, which could be influenced by the Christmas holiday. These observations indicate that holiday periods can influence on EV usage and connectivity patterns, resulting in reduced demand for charging services during these festive times. The analysis at a weekly resolution offers valuable insights into the seasonal fluctuations and the impact of holidays on the behavior of connected EVs throughout the year. It is worth noting that the standard deviation plot aligns with the main plot, exhibiting consistency over the entire year. This consistency suggests that the variability in the number of connected EVs remains relatively stable, with similar levels of dispersion observed throughout the year.

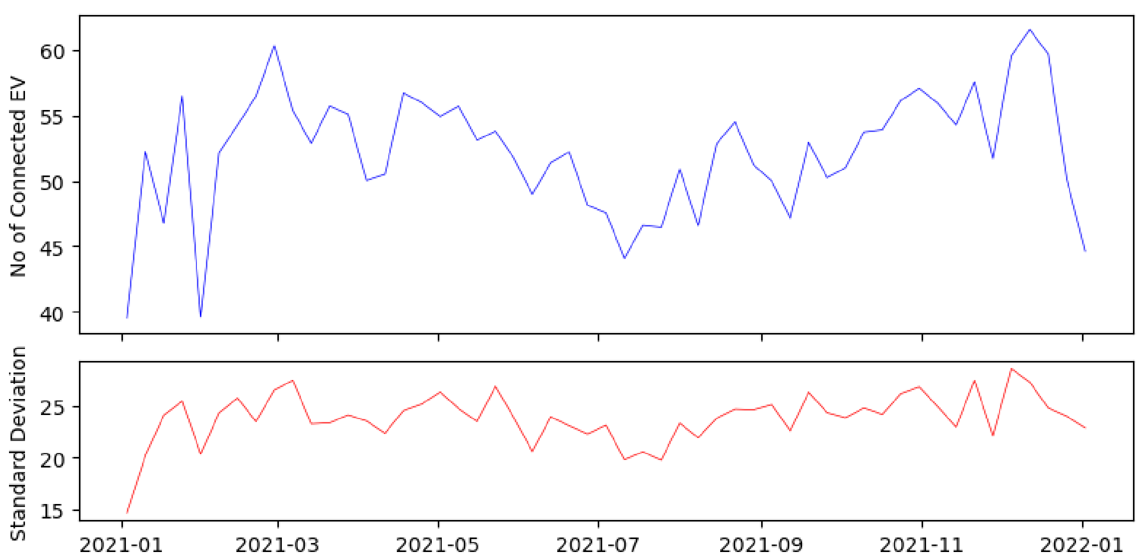

Figure 6-3: Aggregate No. of connected EV along with standard deviation each week.

Further zooming out to the daily plot in Figure 6-4 reveals intriguing patterns in the aggregate number of connected Electric Vehicles (EVs). Notably, significant downward spikes are observed during holiday days, aligning with the previously mentioned reasons for reduced EV usage and connectivity during festive periods. These holiday-related dips reflect a decreased demand for charging services, likely due to reduced travel or charging activities associated with those specific days. Additionally, the daily plot exhibits high fluctuations within each week, with each day showcasing distinct connectivity levels compared to others within the same week. This variability suggests that EV connectivity is influenced by daily factors or routines that contribute to fluctuations in usage patterns. The standard deviation plot follows a similar trend, reflecting the variability and fluctuation observed in the daily plot. Overall, the daily analysis provides valuable insights into the impact of holidays and the daily dynamics of EV usage.

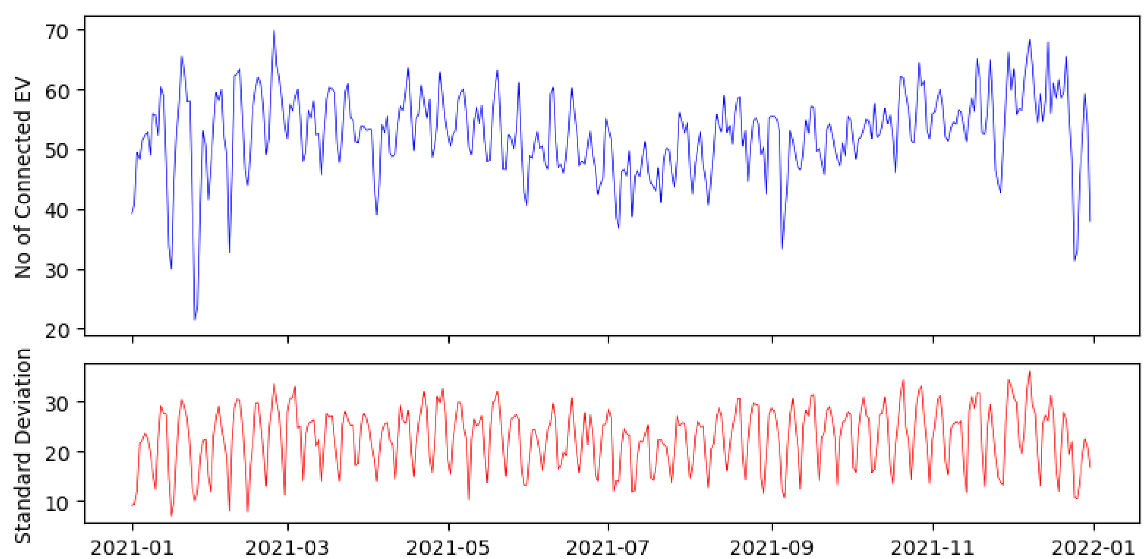

Figure 6-4: Aggregate No. of connected EV along with standard deviation each day.

B. Power

Building upon the previous sections' analysis of the aggregate number of connected Electric Vehicles (EVs), the focus in this section shifts to examining the aggregated power consumed during the charging period of EVs. This analysis aims to provide insights into the overall power consumption behavior and its relationship with EV connectivity.

The following plot in Figure 6-5 showcases the aggregated power consumed per minute throughout the year. Observing the minutes plot, it becomes apparent that the power consumption behavior follows a similar pattern to the EV connectivity trend previously discussed. The stability observed in the EV connectivity plot is reflected in the power consumption plot as well. The power consumed remains relatively consistent over time, with minor variations.

Zooming out to explore the data at different resolutions, including monthly, weekly, and daily intervals, will allow for a more detailed examination of the power consumption patterns. By analyzing power consumption at these different resolutions, potential variations and trends can be uncovered, shedding light on the power consumption behavior of EV users during different time intervals.

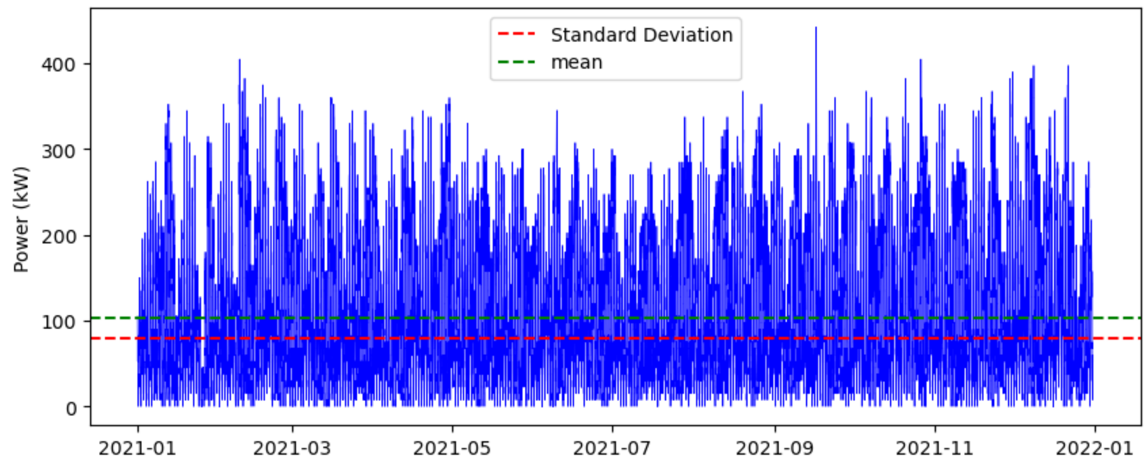

Figure 6-5: Aggregate power consumed during charging per minute.

Examining the plot at a monthly resolution in Figure 6-6, notable patterns emerge. Similar to the previous observations in the EV connectivity analysis, there is a clear trend of higher power consumption during the winter months compared to the summer months. This trend suggests a correlation between weather conditions and power demand, with EV users likely relying more on charging services during colder seasons. However, it is essential to note that within this trend, there are variations and fluctuations in power consumption from month to month. The standard deviation plot, which accompanies the aggregated power plot, helps visualize this variability. It demonstrates the level of dispersion or deviation from the mean power consumption within each month. The standard deviation plot follows a similar pattern to the main plot, indicating that the variability in power consumption is consistent over the year. Additionally, specific months exhibit deviations from the overall trend. For example, January 2021, as previously mentioned, has lower power consumption due to factors like holidays or a potential influx of new EVs joining the network towards the end of the month.

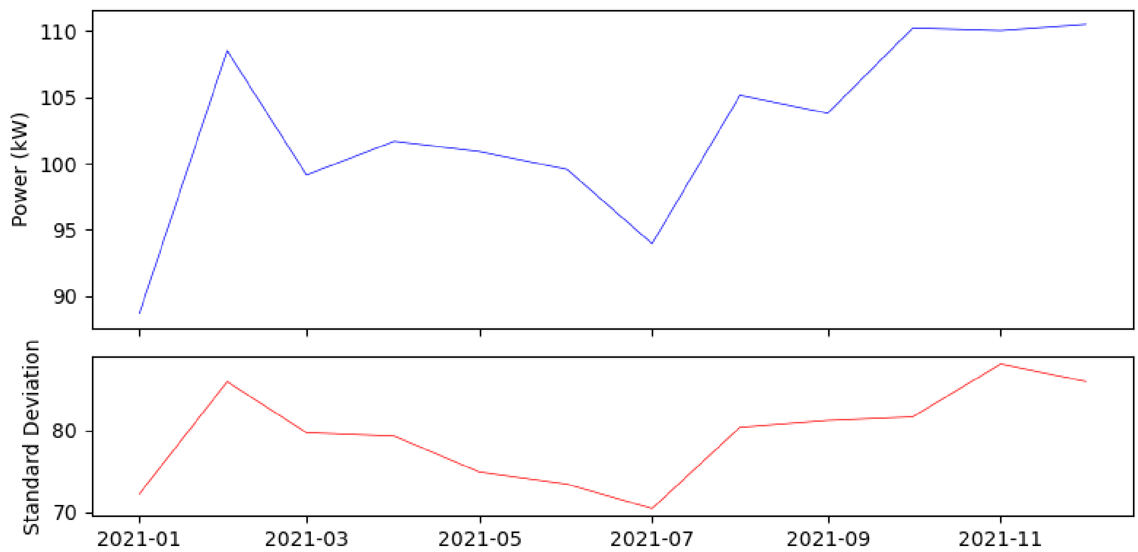

Figure 6-6: Aggregate power consumed during charging along with standard deviation per month.

Similar to the monthly analysis, Figure 6-7 shows that there is a clear trend of higher power consumption during the winter weeks compared to the summer weeks. This observation suggests a relationship between weather conditions and power demand, with EV users likely requiring more charging during colder periods. However, within this overarching trend, there are significant fluctuations in power consumption from week to week. Each week exhibits distinct power consumption levels, indicating variations in EV charging behavior. The standard deviation plot, accompanying the aggregated power plot, highlights this variability. It showcases the level of dispersion or deviation from the mean power consumption within each week. The standard deviation plot follows a similar pattern to the main plot, emphasizing the consistency of power consumption fluctuations throughout the year.

Furthermore, specific weeks deviate from the overall trend. For example, the first week of the year 2021, as mentioned previously, has lower power consumption due to factors like holidays or reduced EV usage during the New Year period. Similarly, the last week of the year, potentially influenced by the Christmas holidays, also exhibits lower power consumption.

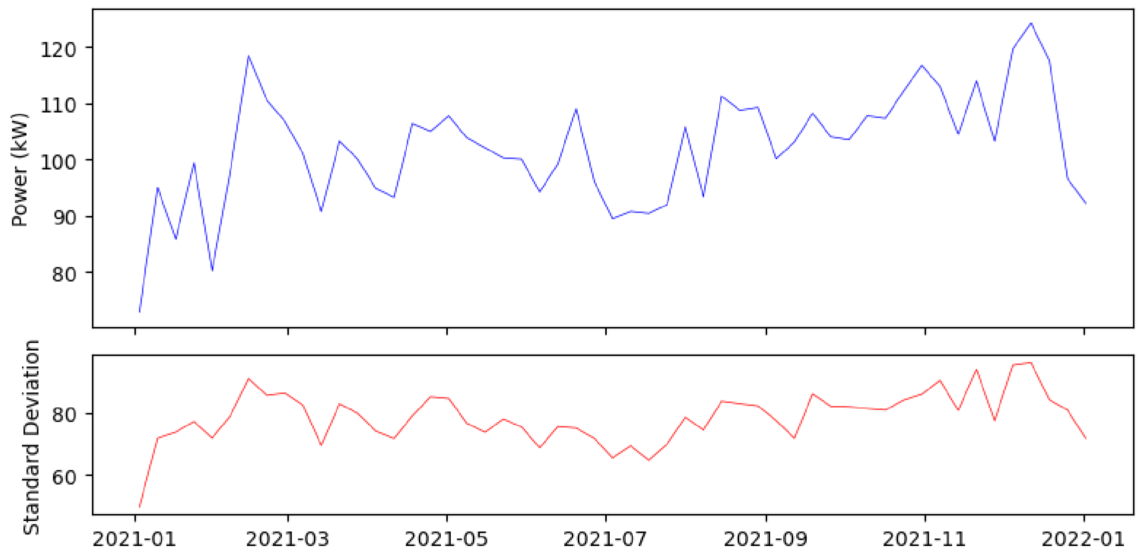

Figure 6-7: Aggregate power consumed during charging along with standard deviation per week.

Examining the daily plot depicted in Figure 6-8 reveals intriguing patterns. A prominent observation is the substantial fluctuations in power consumption observed on daily. Each day showcases unique power consumption levels, underscoring the inherent variability in EV charging behavior. This variability can be attributed to various factors, including individual driving patterns, charging habits, and specific events or circumstances that impact the charging needs of EV users. These factors contribute to the dynamic nature of power demand within the EV charging ecosystem, highlighting the importance of understanding and adapting to the diverse charging patterns of EV owners.

Interestingly, the daily plot also reflects the impact of holidays and festive periods. During holiday days, there is often a visible dip in power consumption, indicating a reduced demand for charging services. This can be attributed to factors like reduced travel, fewer

EVs in use, or altered routines during those specific days. The standard deviation plot, accompanying the aggregated power plot, demonstrates the level of dispersion or deviation from the mean power consumption within each day. It highlights the daily fluctuations in power consumption and provides insights into the range of power demand experienced throughout the year.

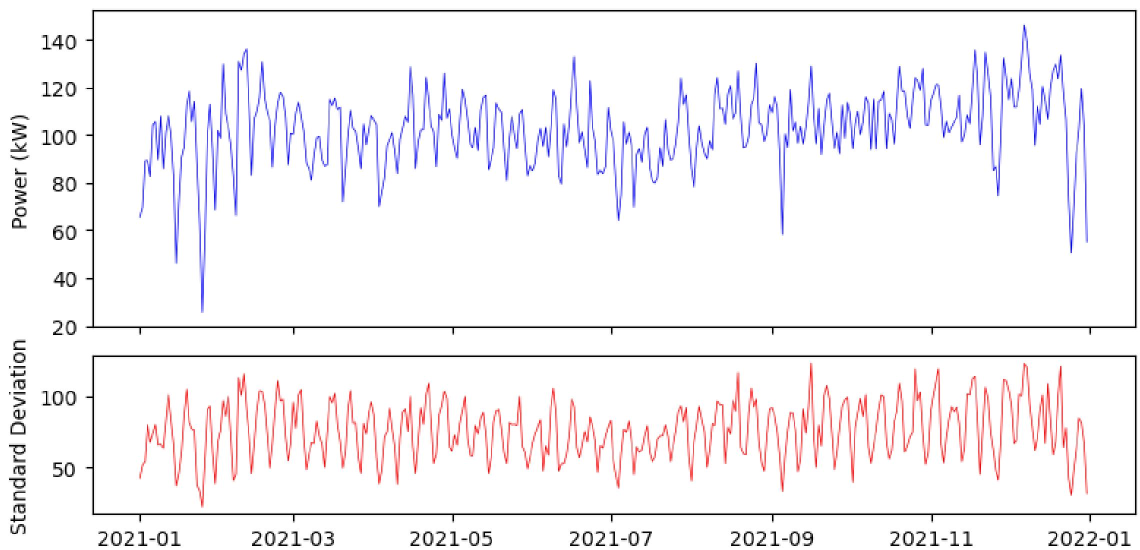

Figure 6-8: Aggregate power consumed during charging along with standard deviation per day.

6.2.2 Averaging EV User Behavior Over Time

6.2.2.1 Introduction

In examining the data on EV connectivity and power consumption, a consistent pattern emerges on a monthly, weekly, and daily basis. However, there are noticeable differences between different periods. To gain a comprehensive understanding of EV user behavior over time, it is necessary to analyze the data by averaging the values for specific periods within a month, week, and day. The process begins by calculating the mean values for each day of the month. By aggregating the data across multiple months, the typical behavior exhibited by EV users on each specific day can be identified. Moving on to the weekly average, the data is aggregated for each day of the week across multiple weeks. This allows for the identification of patterns and tendencies specific to each day of the week in terms of EV connectivity and power consumption. Furthermore, the analysis focuses on the time of day within a 24 hours. By aggregating the data for each specific hour of the day across multiple days, the mean values for EV connectivity and power consumption can be determined. This provides insights into the typical behavior and usage patterns of EV users at different times throughout the day. By employing this process of averaging, a more comprehensive understanding of EV user behavior over time can be achieved. This analysis offers valuable insights into the typical mean values of EV connectivity and power consumption for specific periods within a month, week, and day.

6.2.2.2 Results

A. No. of Simultaneously Connected EVS

In Figure 6-9, the EV connectivity data is visualized over a year, showing the relationship between the day of the month and the number of connected EVs. To determine the typical mean for each day, the data is aggregated across all 12 months. By summing the values from all months, the total count of connected EVs is obtained for each specific day. Dividing this count by 12 gives the average number of connected EVs for that day, representing the typical mean across all months. In addition to the mean, the standard deviation is calculated to measure the variability in EV connectivity for each day.

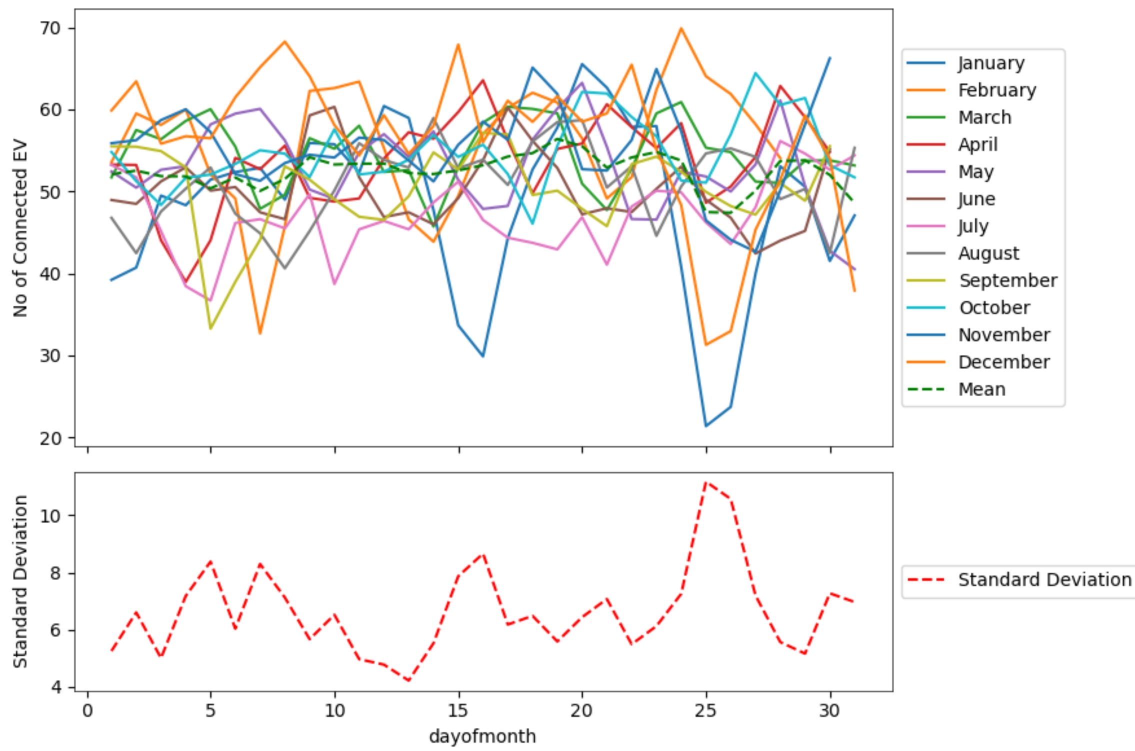

Figure 6-9: Monthly variation in No. of connected EV along with standard deviation.

Figure 6-10 presents the mean and standard deviation for each day of the month, providing valuable insights into the typical patterns of EV connectivity throughout the year. Notably, a significant decline is observed on the 25th day, deviating from the overall trend. This observation may be influenced by the behavior observed in January and February, which likely contributed to a decrease in the mean during that period. However, apart from this deviation, the mean and standard deviation for the remaining days demonstrate relatively consistent values. The mean values represent the average level of EV connectivity for each day of the month, while the standard deviation indicates the degree of variation around the mean. The overall consistency in the mean and standard deviation suggests a stable pattern in EV connectivity, with minimal fluctuations or irregularities for the majority of the year.

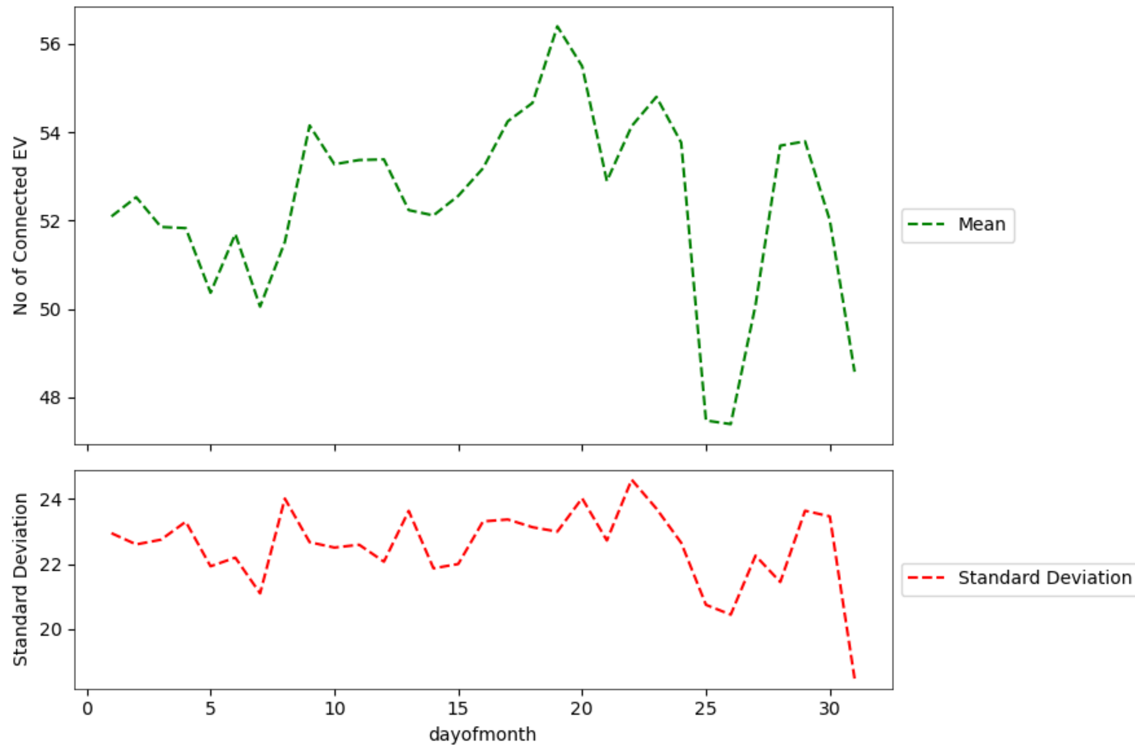

Figure 6-10: Monthly variation in average No. of connected EVs along with standard deviation.

Figure 6-11 illustrates the mean and standard deviation for each day of the week, highlighting distinct variations in the mean values across different days. Upon examining the plot, it becomes apparent that each day of the week presents a unique level of EV connectivity. Specifically, midweek days consistently demonstrate higher levels of EV connectivity, whereas Fridays and weekends typically exhibit lower levels. These patterns can be attributed to factors such as commuting patterns, discrepancies in EV usage between weekdays and weekends, or specific events and conditions that influence EV utilization.

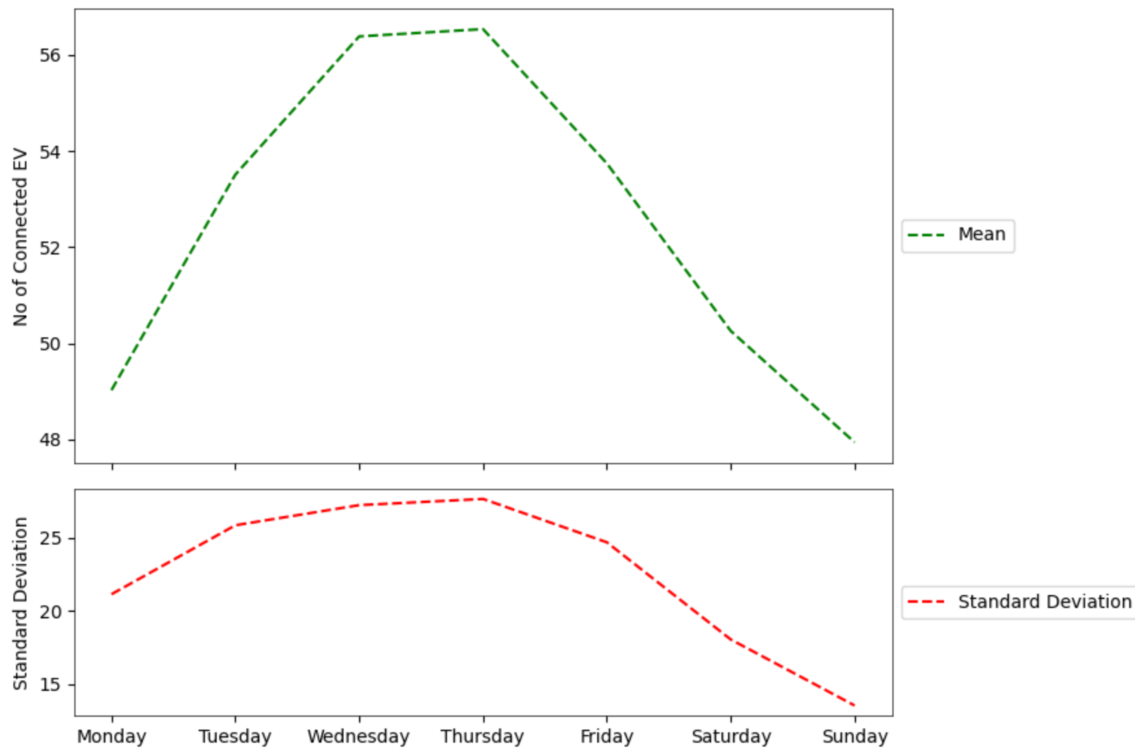

Figure 6-11: Weekly variations in average No. of connected EVs along with standard deviation.

Figure 6-12 displays the mean and standard deviation for each hour of the day, offering valuable insights into the typical characteristics of EV connectivity throughout a 24 hours period. Notable patterns emerge, indicating specific hours when EV connectivity consistently reaches higher levels, signifying peak periods of usage. These peak hours, typically occurring in the evening and late-night, align with times when users are likely to be at home and have their EVs connected. Conversely, there are hours when EV connectivity is lower, typically coinciding with periods when individuals are away from home, such as during work hours or other activities. The observed variations in EV connectivity throughout the day can be attributed to various factors, including commuting patterns, work schedules, and individual charging behaviors. Furthermore, it is worth mentioning that the standard deviation shows an intriguing pattern with higher values around 7 am to 8 am, which aligns with the time when users typically leave for work. This observation indicates a potential variation in EV charging behaviors during the morning rush hour period, implying that there may be diverse charging patterns or preferences among EV users during this time.

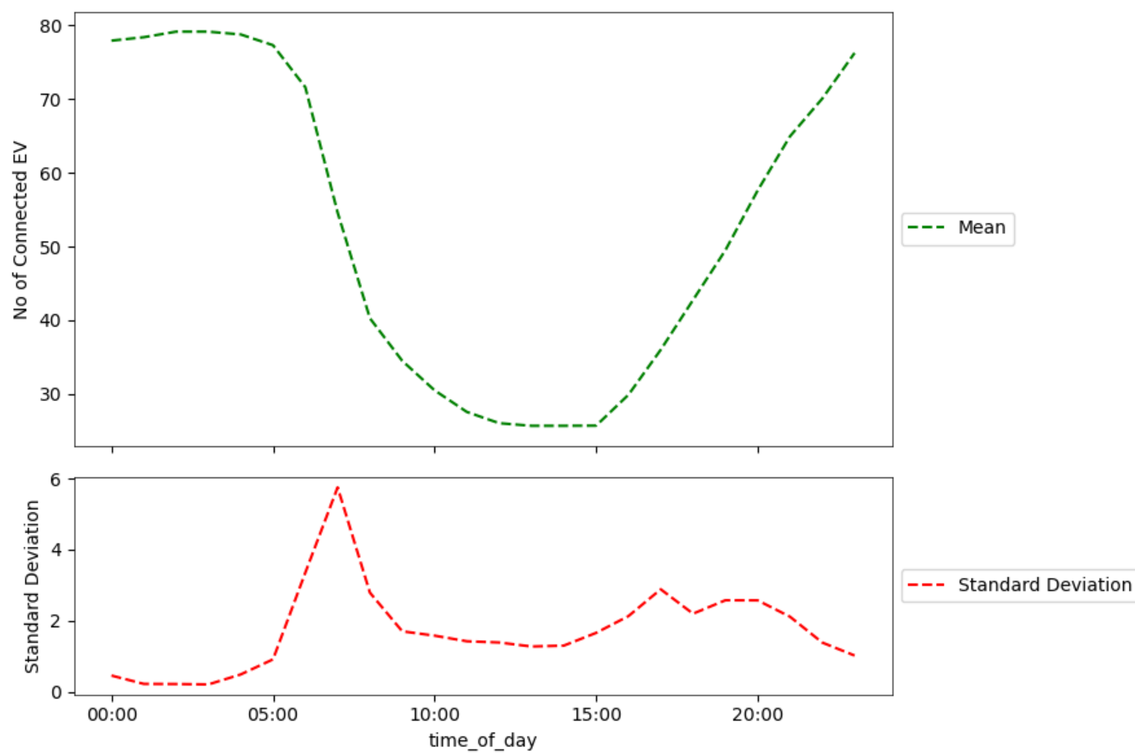

Figure 6-12: Daily variations in average No. of connected EVs along with standard deviation.

B. Power

Shifting the focus to the power consumption behavior, Figure 6-13 offers valuable insights into the typical power consumption patterns of EVs throughout the month. It is worth noting that there are some variations in power consumption levels on different days of the month. Some days may display higher power consumption, indicating increased charging activity, while others, like day 25, consistently exhibit lower values. Additionally, it is important to consider that day 31 may not exist every month, which can also contribute to variations in power consumption patterns. These variations in power consumption can be influenced by factors such as individual driving habits, charging preferences, and specific events or circumstances that impact EV usage.

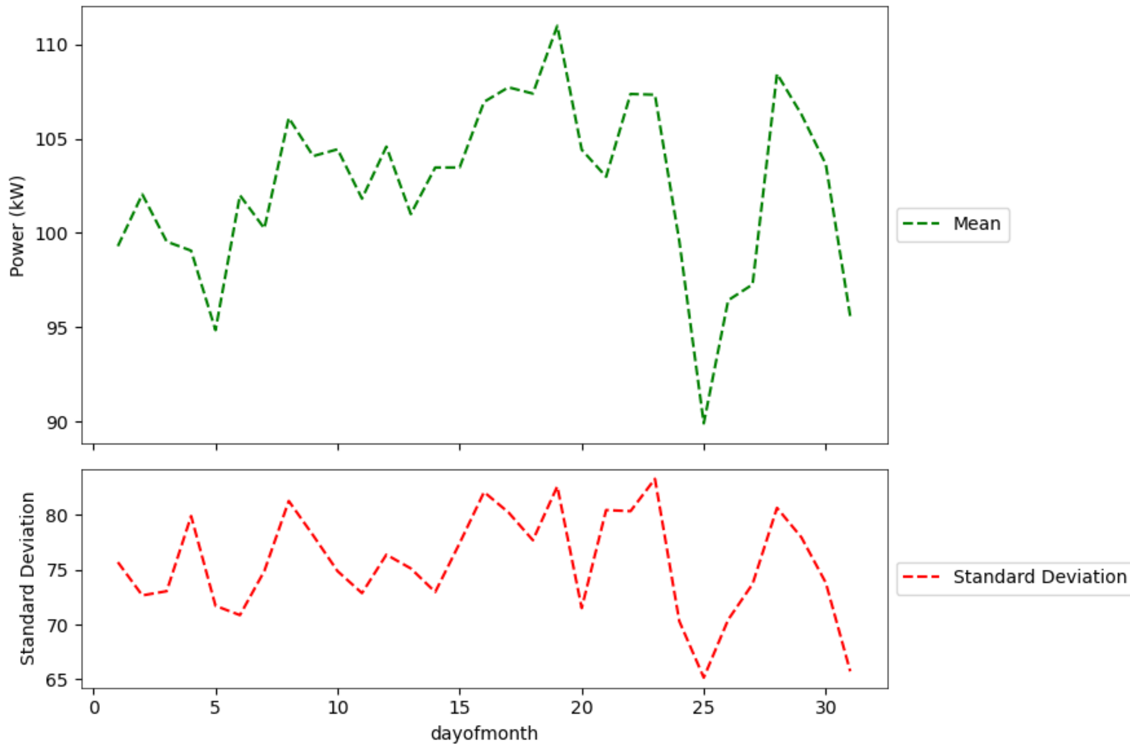

Figure 6-13: Monthly variations in power consumed during charging along with standard deviation.

Figure 6-14 offers valuable insights into the typical power consumption behavior of EVs throughout the week. Weekdays, corresponding to workdays, tend to exhibit higher power consumption due to increased charging activity driven by commuting patterns and regular work schedules. In contrast, Fridays and weekends show lower power consumption as fewer individuals use their EVs for work-related purposes during these days. These variations in power consumption can be attributed to factors such as diverse usage patterns, charging behaviors, and individual routines. Additionally, it is important to note the significant difference in the standard deviation between weekends and workdays, indicating varying charging needs and behaviors of EV users throughout the week.

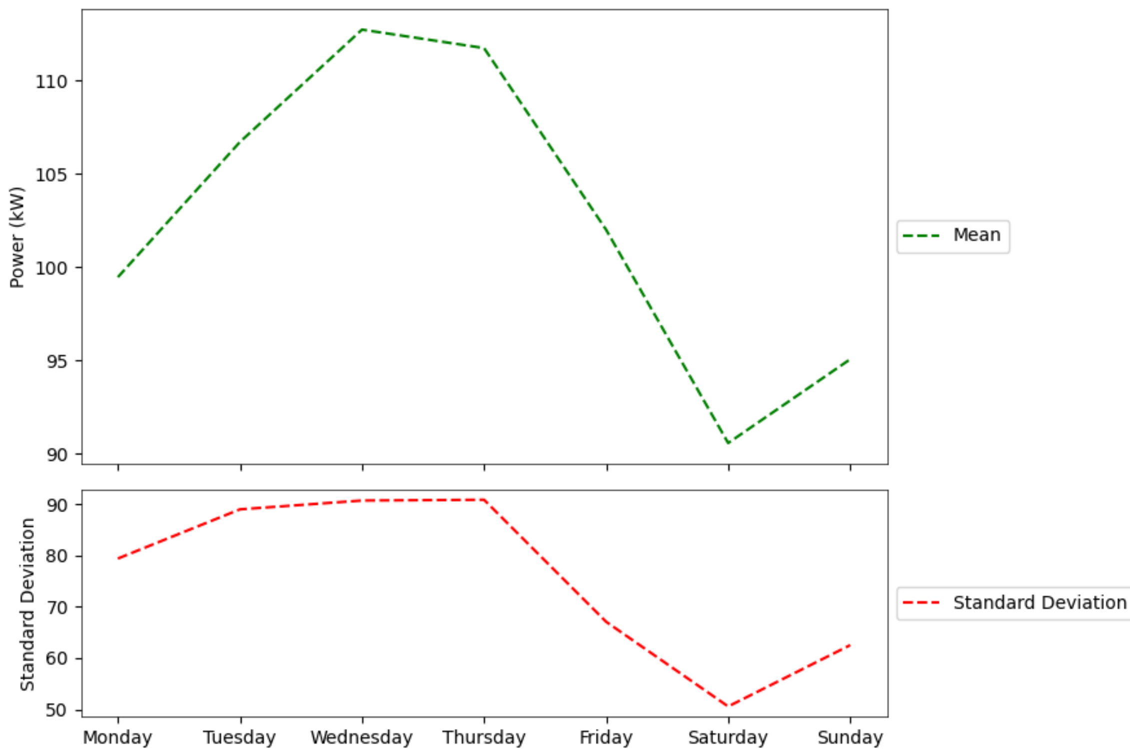

Figure 6-14: Weekly variations in power consumed during charging along with standard deviation.

Figure 6-15 provides valuable insights into the typical power consumption behavior of EVs over a 24 hours period. There are specific hours when power consumption consistently peaks, indicating periods of heightened charging activity. These peak hours typically occur in the evening and late at night, between 6 pm and 12 am when EV users are more likely to be at home. Conversely, power consumption is lower during hours when individuals are away from home, such as during work hours or other daily activities. Additionally, power consumption tends to be lower in the early morning when EVs are already fully charged.

The standard deviation plot complements this information by showcasing the variability in power consumption throughout the day. It reveals higher levels of variability in the evening hours when users arrive home from work and engage in charging activities. In contrast, the morning and afternoon hours exhibit lower variability, indicating more consistent power consumption patterns during those times.

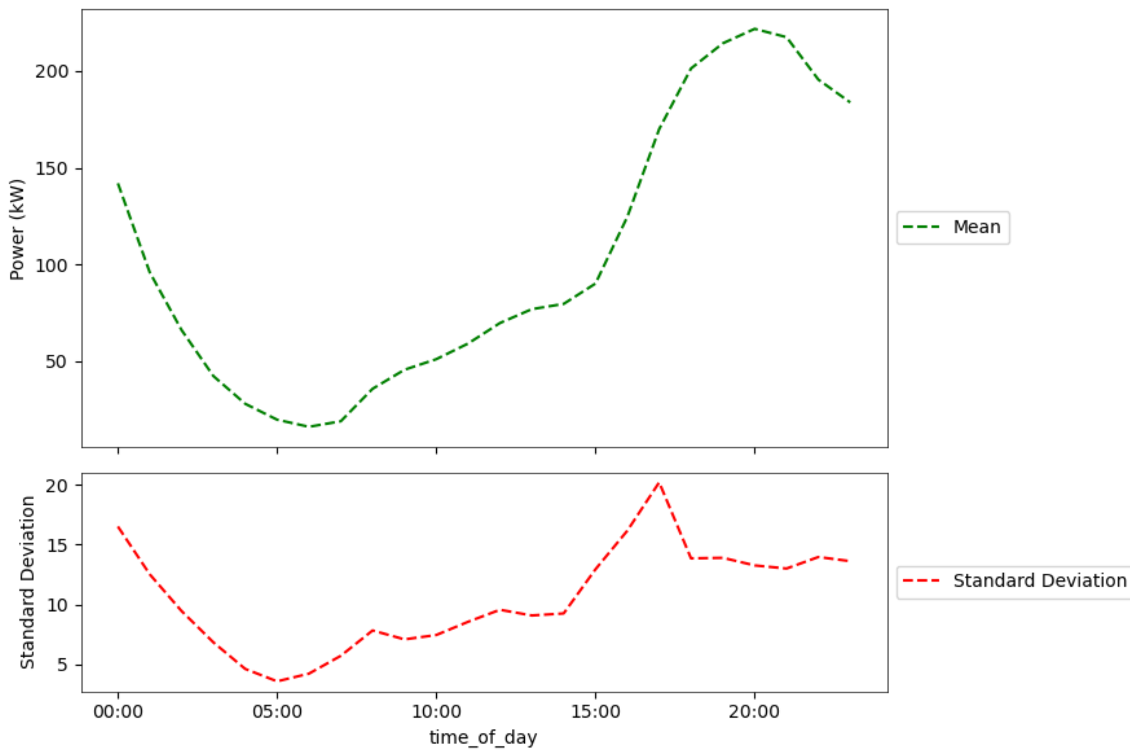

Figure 6-15: Daily variations in power consumed during charging along with standard deviation.

6.2.3 Machine Learning Prediction of Aggregated EV User

6.2.3.1 Introduction

This section focuses on evaluating the effectiveness of using the sessions data model to predict aggregated charging behavior. Rather than relying on machine learning or time series techniques, the approach involves leveraging the sessions data model and utilizing its predictions in the analysis of aggregated charging behavior.

The sessions data model captures individual charging sessions and their characteristics, allowing for the aggregation of predicted charging behavior to estimate overall charging activity. This estimation takes into account the predicted behavior of multiple charging session, providing insights into broader charging patterns and trends.

To assess the accuracy of the predictions, a comparison is conducted between the estimated aggregated charging behavior and the actual aggregated behavior observed in the dataset. By examining the disparities between the predicted and actual values, the performance of the sessions data model in capturing collective charging behavior can be evaluated.

This comparative analysis offers insights into the effectiveness of the sessions data model in predicting aggregated charging behavior. It helps to understand the model's ability to capture the inherent variability and patterns within the dataset, offering valuable information for future predictions and optimizations in EV charging infrastructures.

6.2.3.2 Results

A. No. of Simultaneously Connected EVs

Figure 6-16 presents a visual comparison of the actual and predicted number of connected EVs per minute throughout the year, allowing for an assessment of the model's performance in forecasting charging behavior. The Root Mean Square Error (RMSE), calculated at 12.3, provides a quantitative measure of the disparity between the predicted and actual values, representing the overall error in the model's predictions. A lower RMSE signifies a higher degree of agreement between the predicted and actual charging behavior, indicating a more accurate model. Apart from RMSE, additional statistical measures, namely the Mean Absolute Error (MAE) and Mean Percentage Error (MPE), play a significant role in evaluating the model's performance and accuracy. The MAE, computed as 9.9, represents the average magnitude of errors between the predicted and actual values. It provides insight into the average deviation between the two sets of values. On the other hand, the MPE, with a value of -7.8%, indicates the average percentage deviation between the predicted and actual values. The negative sign suggests that, on average, the predicted values are lower than the actual values.

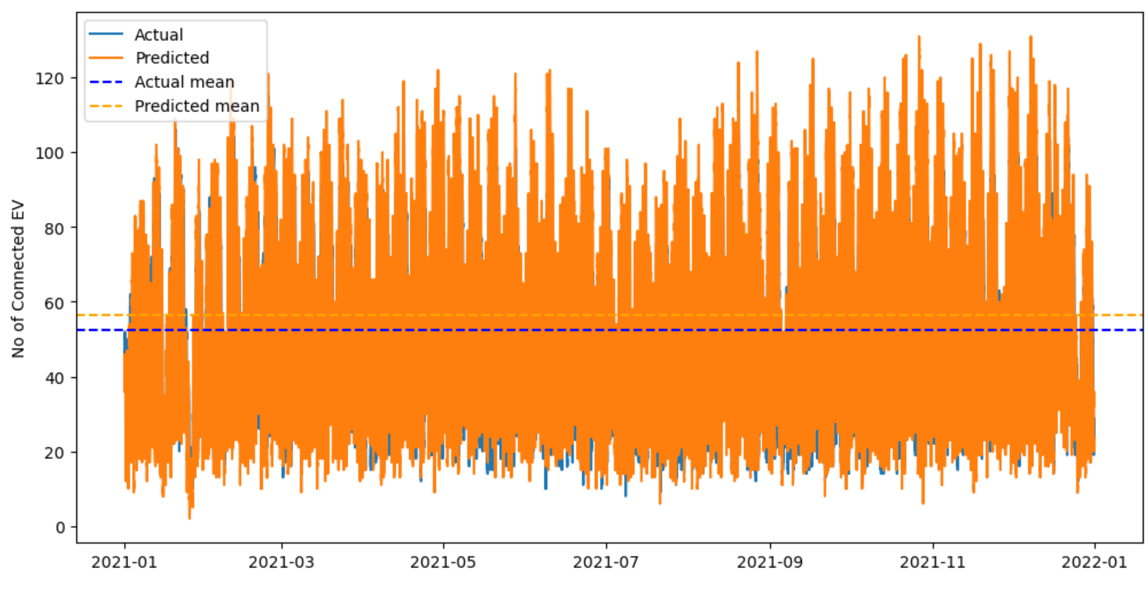

Figure 6-16: Actual vs predicted aggregate No. of EV connected per minute.

Figure 6-17 depicts the hourly actual and predicted number of connected EVs throughout the year, enabling an evaluation of the model's performance in forecasting charging behavior at an hourly resolution. This comprehensive overview allows for a thorough assessment of the model's accuracy and effectiveness. The Root Mean Square Error (RMSE) is calculated to be 12.3, indicating the level of discrepancy between the predicted and actual values. Moreover, the Mean Absolute Error (MAE) is computed as 9.3, while the Mean Percentage Error (MPE) is determined to be -7.8%

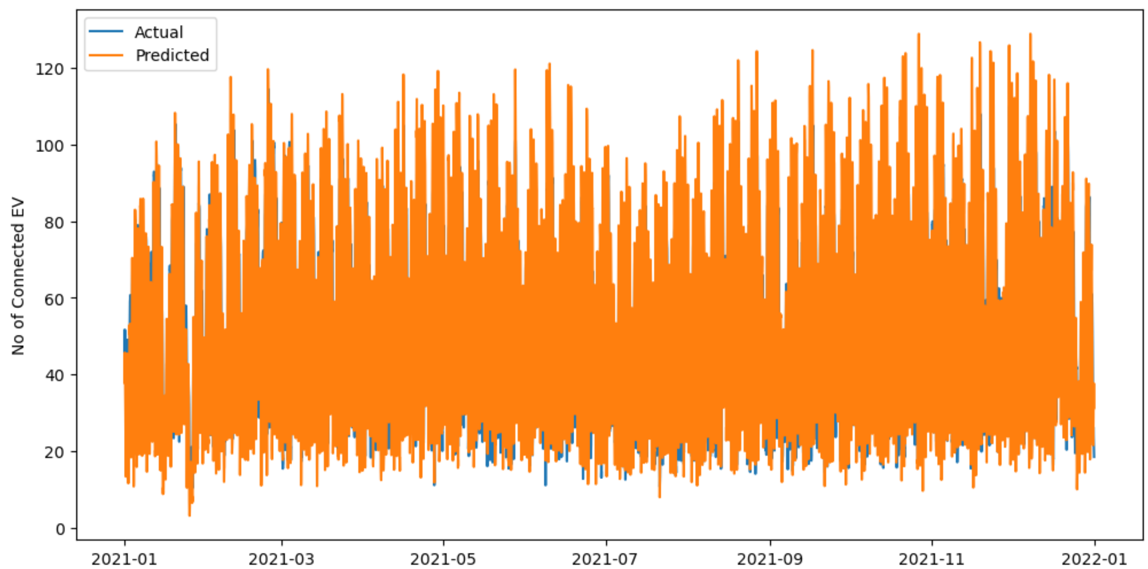

Figure 6-17: Actual vs predicted aggregate No. of EV connected per hour.

At a daily level, Figure 6-18 presents a daily comparison between the actual and predicted number of connected EVs throughout the year, facilitating a comprehensive assessment of the model's forecasting performance in charging behavior. The statistical metrics reveal the model's accuracy, with an RMSE value of 5.8, an MAE value of 4.8, and an MPE value of -7.4%. These metrics provide valuable insights into the average level of error and the extent of deviation between the predicted and actual values, aiding in the evaluation of the model's predictive capabilities at a daily level.

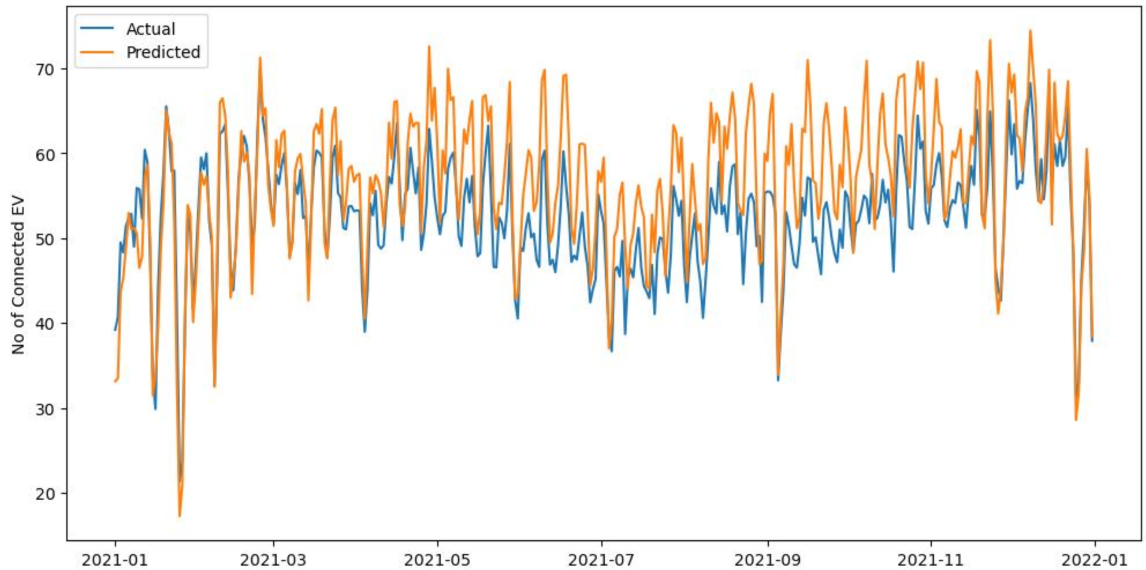

Figure 6-18: Actual vs predicted aggregate No. of EV connected per day.

Expanding the analysis to a weekly timeframe, Figure 6-19 showcases the actual and predicted number of connected EVs per week throughout the year, offering a holistic evaluation of the model's forecasting accuracy in charging behavior. Assessing the model's performance at the weekly level involves calculating the Root Mean Square Error (RMSE), which results in a value of 5, and the Mean Absolute Error (MAE), with a value of 4.4.

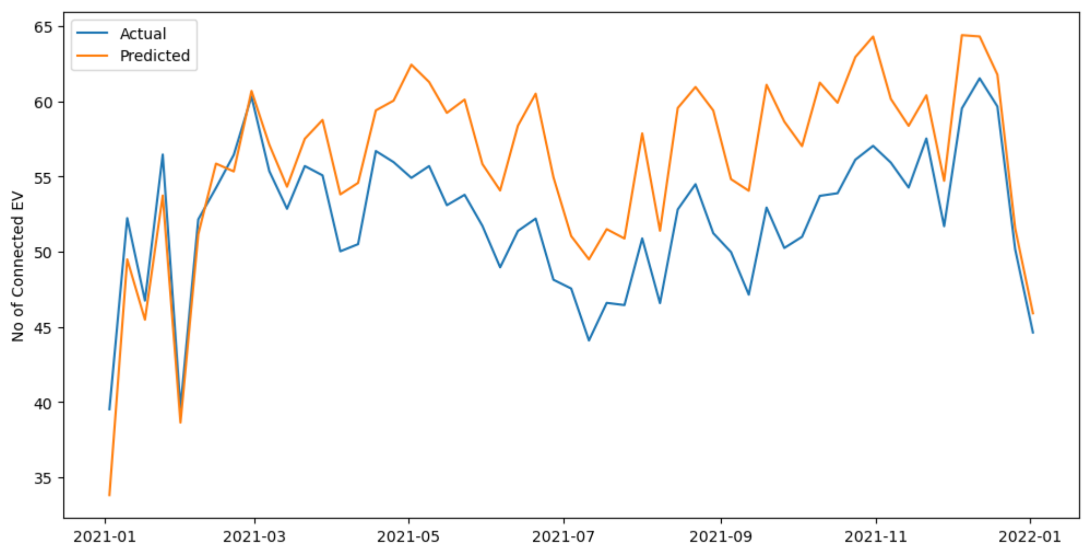

Figure 6-19: Actual vs predicted aggregate No. of EV connected per week.

At a monthly level, Figure 6-20 provides a comparison between the actual and predicted number of connected EVs for each month throughout the year. This expanded perspective allows for a comprehensive evaluation of the model's accuracy in forecasting charging behavior every month. The Root Mean Square Error (RMSE) value for this model is determined to be 4.8, indicating the overall deviation between the predicted and actual values. The Mean Absolute Error (MAE) is calculated as 4.3, representing the average magnitude of errors between the predicted and actual values. Additionally, the Mean Percentage Error (MPE) is calculated as -7.4%, reflecting the average percentage deviation between the predicted and actual values.

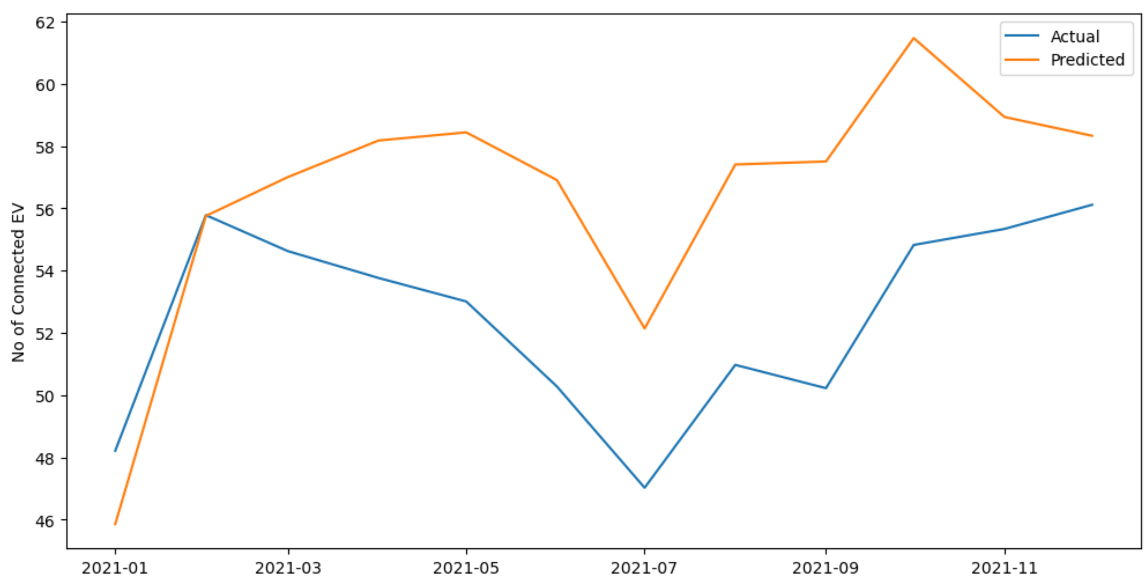

Figure 6-20: Actual vs predicted aggregate No. of EV connected per month.

B. Power

When examining power consumption at a minute-level resolution, Figure 6-21 provides valuable insights into the behavior of EVs. It reveals that certain minutes may experience higher power consumption, indicating increased charging activity or a higher demand for power. Conversely, there are minutes when power consumption is lower, indicating reduced charging activity or lower demand for power. In this case, the RMSE is measured at 20.8, the MAE at 15.2, and the MPE at xx%. These metrics enable an evaluation of the accuracy and reliability of the power consumption predictions at a minute-level granularity.

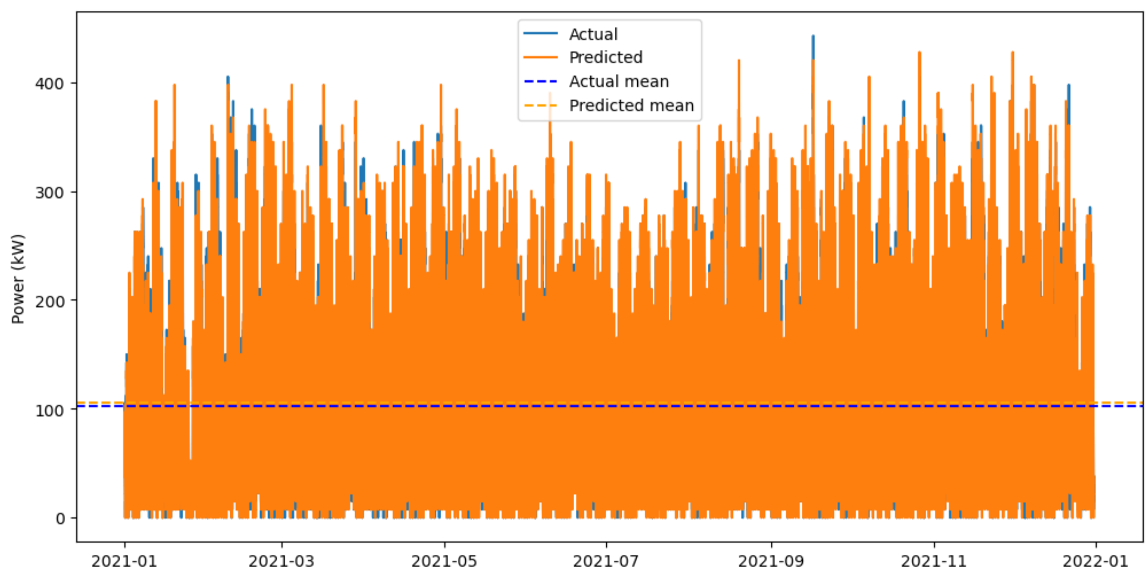

Figure 6-21: Actual vs predicted aggregate power consumed during charging per minute.

Examining the power consumption behavior of EVs at an hourly resolution offers a comprehensive view of the trends and patterns over time. Examining Figure 6-22 allows for the identification of specific hours with higher power consumption, indicating increased charging activity or greater power demand. Conversely, there may be hours with lower power consumption, suggesting reduced charging activity or lower power demand during those periods. The Root Mean Square Error (RMSE) is calculated to be 20.8, the Mean Absolute Error (MAE) is 13.8, and the Mean Percentage Error (MPE) is x%, offering insights into the accuracy and deviation between the predicted and actual power consumption values at an hourly level.

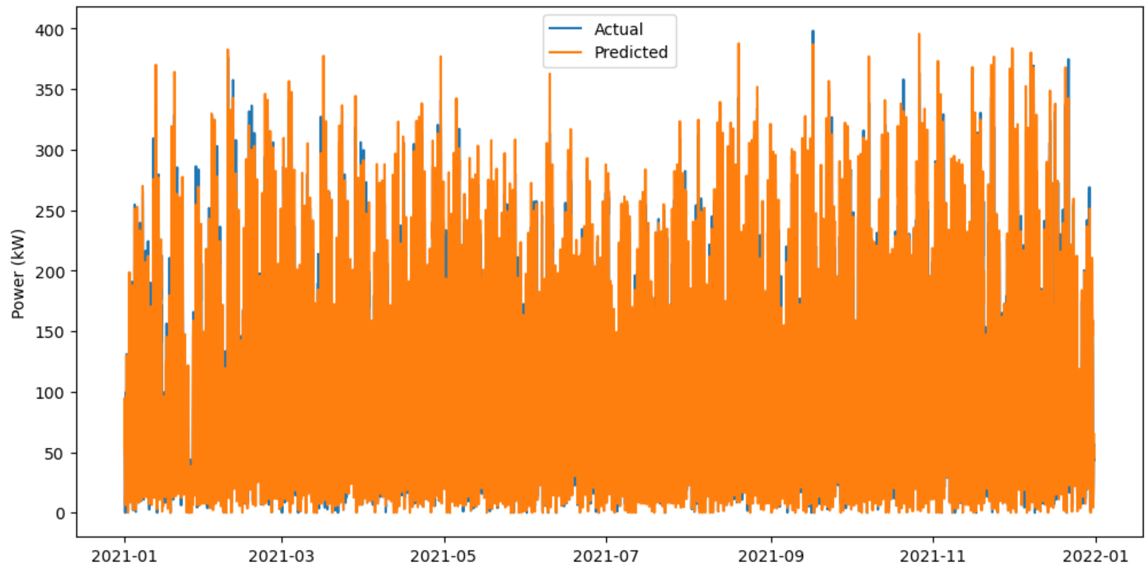

Figure 6-22: Actual vs predicted aggregate power consumed during charging per hour.

Expanding the analysis to a daily resolution, Figure 6-23 provides a broader perspective on the power consumption behavior of EVs over time. By examining the plot, specific days with higher power consumption and days with lower power consumption can be identified. The Root Mean Square Error (RMSE) is calculated to be 6.7, the Mean Absolute Error (MAE) is 5.4, and the Mean Percentage Error (MPE) is -3.5%. These statistical metrics offer insights into the accuracy and deviation between the predicted and actual power consumption values at a daily level. It is worth noting that the predicted trend generally follows the same pattern as the actual trend, albeit with higher values. This suggests that the model captures the overall behavior of power consumption, albeit with some deviations in the magnitude of the values

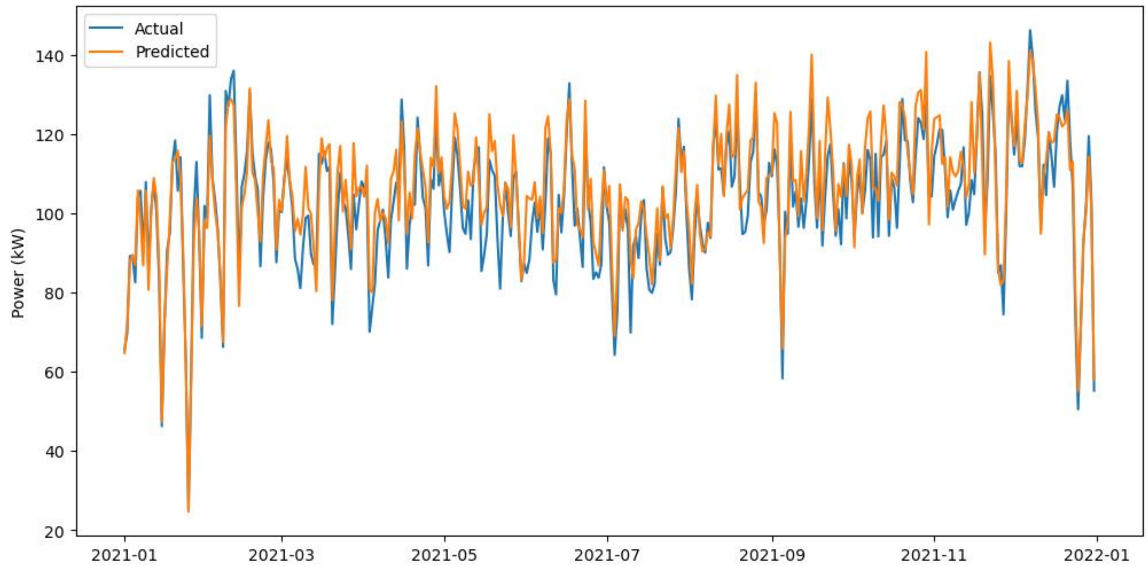

Figure 6-23: Actual vs predicted aggregate power consumed during charging per day.

Zooming out to a weekly resolution, a comprehensive understanding of the power consumption behavior of EVs over time can be obtained. By examining Figure 6-24, which represents the aggregated power consumption over the weeks of the year, insights into the trends and patterns of power usage on a weekly basis can be observed. The Root Mean Square Error (RMSE) is determined to be 4.4, indicating the average deviation between the predicted and actual power consumption values. The Mean Absolute Error (MAE) is calculated as 3.9, representing the average magnitude of the errors. Furthermore, the Mean Percentage Error (MPE) is calculated as -3.2%, indicating the average percentage deviation between the predicted and actual values. It is worth noting that the predicted trend generally follows the same pattern as the actual trend, albeit with higher values. This suggests that the model captures the overall behavior of power consumption, albeit with some deviations in the magnitude of the values

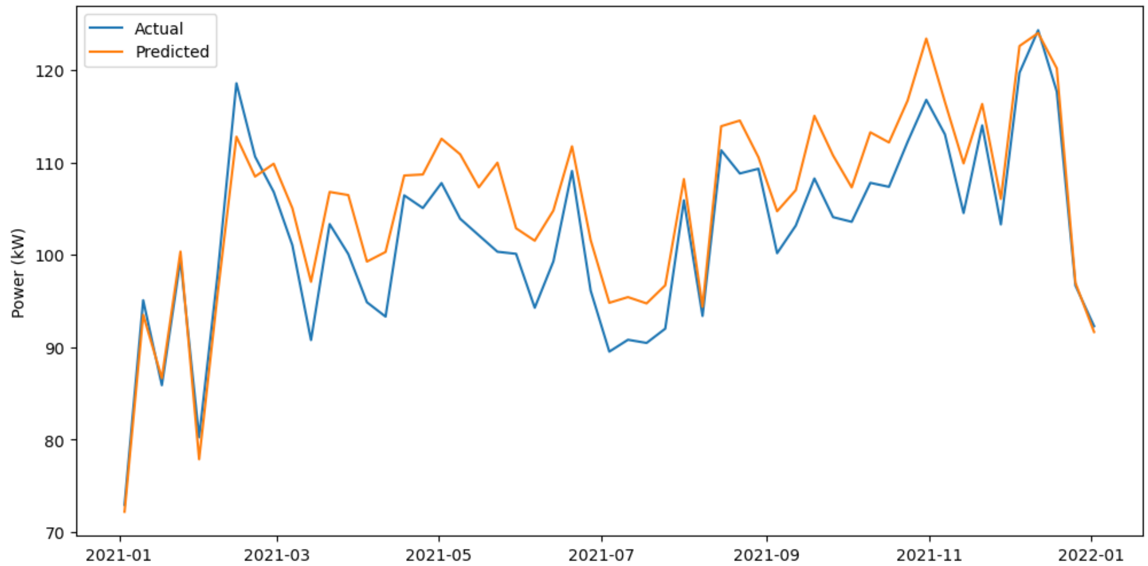

Figure 6-24: Actual vs predicted aggregate power consumed during charging per week.

When analyzing the power consumption behavior of EVs at a monthly resolution, a comprehensive understanding of their charging patterns throughout the year can be obtained. Figure 6-25 enables the identification of months with higher power consumption, indicating increased charging activity or higher power demand. Conversely, some months exhibit lower power consumption, suggesting reduced charging activity or lower power demand during those periods. It is noteworthy that the predicted trend generally aligns with the actual trend, albeit with higher values. The Root Mean Square Error (RMSE) is determined to be 4.1, representing the average discrepancy between the predicted and actual power consumption values. The Mean Absolute Error (MAE) is calculated as 3.7, indicating the average magnitude of the errors. Additionally, the Mean Percentage Error (MPE) is calculated as -3.2%, reflecting the average percentage deviation between the predicted and actual values.

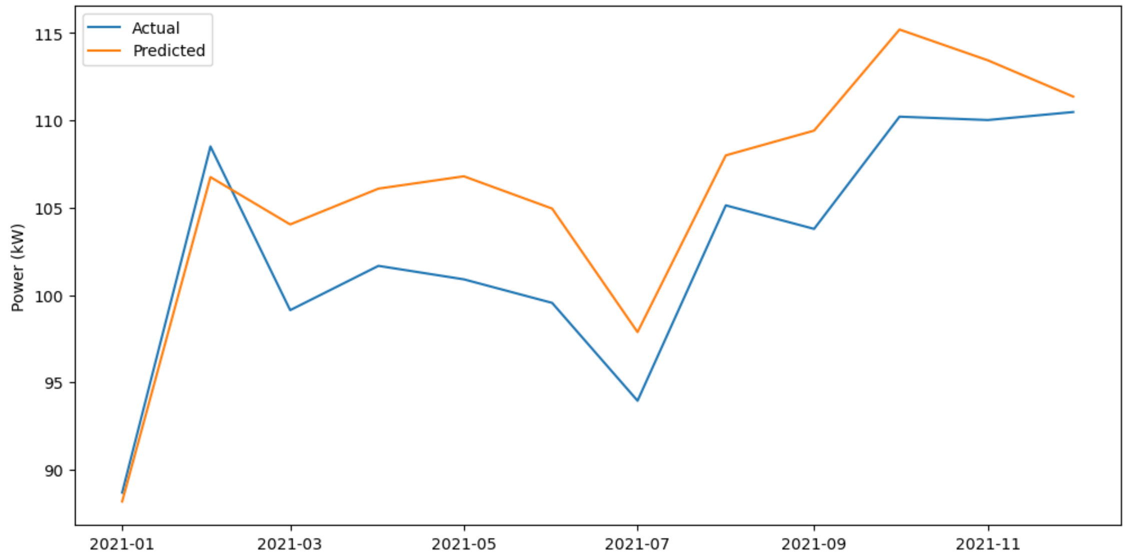

Figure 6-25: Actual vs predicted aggregate power consumed during charging per month.

6.3 Conclusions

In conclusion, the analysis of aggregated EV connectivity data in 2021 reveals stable and consistent EV usage patterns throughout the year. Winter months show higher EV connectivity, while holiday periods exhibit lower demand for charging services. Daily fluctuations indicate the influence of daily routines on EV usage. Similarly, the analysis of aggregated power consumption during EV charging reveals consistent patterns and trends. Higher power consumption is observed during winter months, indicating a correlation with weather conditions. Fluctuations in power consumption occur at monthly, weekly, and daily intervals, reflecting variations in EV charging behavior. Holiday periods show reduced power demand, while daily fluctuations highlight the dynamic nature of power demand influenced by individual habits and circumstances. Understanding these patterns is crucial for effective infrastructure planning and accommodating diverse charging needs.

Analyzing the data through averaging provides a more comprehensive understanding of EV user behavior over time. By calculating mean values for specific periods within a month, week, and day, patterns and tendencies in EV connectivity and power consumption can be identified. The analysis reveals consistent behavior for most days of the month, with a noticeable decline on the 25th day. Weekly analysis shows distinct variations in EV connectivity across different days of the week, with midweek days demonstrating higher levels and Fridays and weekends showing lower levels.

The effectiveness of the sessions data model in predicting aggregated charging behavior is evaluated by comparing the predicted values with the actual values observed in the dataset. This comparison is conducted at different time resolutions, including minutes, hours, days, weeks, and months. The evaluation involves calculating statistical measures such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Percentage Error (MPE) to assess the model's performance and accuracy.

For the prediction of simultaneous connections, the comparison shows that at the minute level, the RMSE is 12.3, indicating the overall error in the model's predictions. The MAE is 9.9, representing the average magnitude of errors, and the MPE is -7.8%, suggesting that the predicted values are, on average, higher than the actual values. At the hourly level, the RMSE is 12.3, the MAE is 9.3, and the MPE is -7.8%.

In conclusion, the predictive model showed varying levels of accuracy at different timescales. It performed reasonably well at weekly and monthly predictions, with reduced error compared to daily and minute-by-minute connections.

7. IMPLICATIONS OF PREDICTED USER BEHAVIOR ON SCHEDULING

7.1 Overview

The previous chapter delved into an analysis of EV user behavior, uncovering key patterns and characteristics that influence charging behavior, including factors like time of day, day of the week, and contextual information. This analysis serves as a solid foundation for the current chapter, which explores the implications of predicted user behavior on scheduling. By leveraging the predictive models developed in the previous chapter, it becomes possible to forecast connection duration and charging duration for EV charging sessions. These predictions have far-reaching implications, enabling optimization of scheduling processes, resource allocation, and capacity planning across different domains.

Conventional scheduling approaches often rely on simplified assumptions or historical data, which may not capture the dynamic and evolving nature of user charging behavior. This limitation hampers the effective management of charging resources, resulting in suboptimal outcomes such as increased peak demand, higher electricity costs, and lower user satisfaction. Consequently, there is an urgent need to investigate innovative methods that can provide accurate predictions of user behavior and integrate them seamlessly into the scheduling process. By doing so, organizations can enhance their scheduling decisions, maximize operational efficiency, improve user experiences, and achieve better resource utilization.

In this section, a single day schedule is developed based on predicted behavior of a group of EV users. This schedule is evaluated against the actual user behavior to assess the satisfaction of charging demand.

7.2 Methodology

This section outlines the methodology used in the study, encompassing both data preparation and analysis stages. In this chapter, a dataset consisting of 140 EV charging sessions was used. The selected sessions were purposefully sampled from a specific day, specifically July 22nd, 2021, which happened to be a Tuesday. This day was chosen due to its representation of the typical charging patterns observed in EV charging activities. Figure 7-1 presents the number of connected EVs on July 22nd, compared to the average of all days in the dataset, showcasing the typical daily pattern.

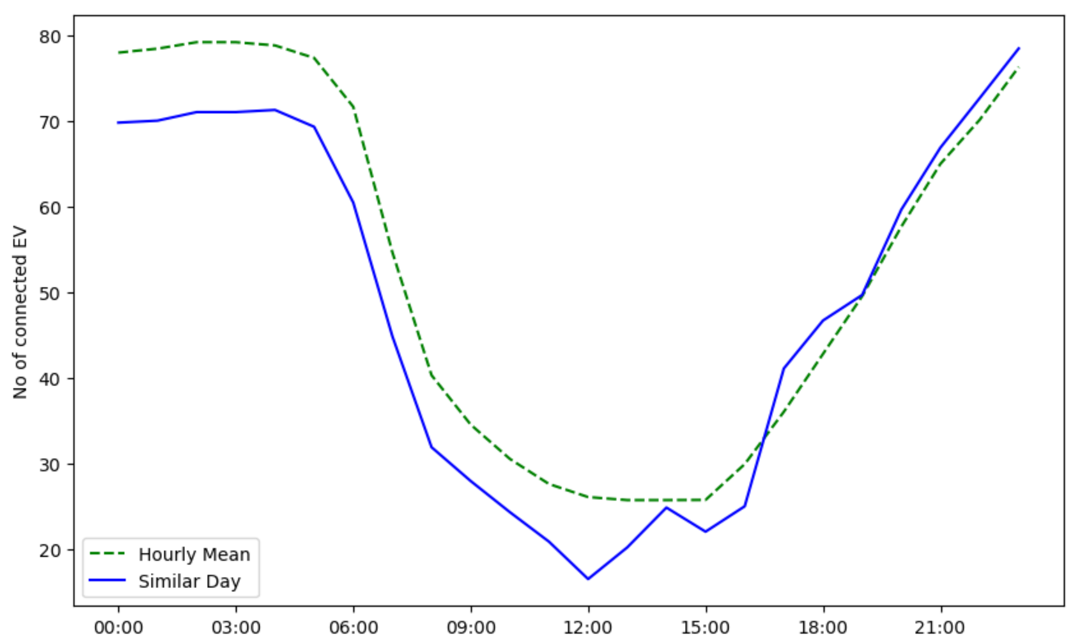

Figure 7-1: Hourly EV connections on July 22, 2021, and average of all days.

For the analysis of EV charging behaviors in a residential context, it was assumed that the charging sessions took place in 114 households equipped with attached garages. Figure 7-2 displays the load profile, providing a visual representation of the electricity usage patterns. This assumption was made to consider factors such as the availability of charging infrastructure and charging patterns specific to households. To simulate the impact of EV penetration in the area, a multiplication factor of 2.2 was applied. This factor represents a 50% penetration rate of EVs in the region.

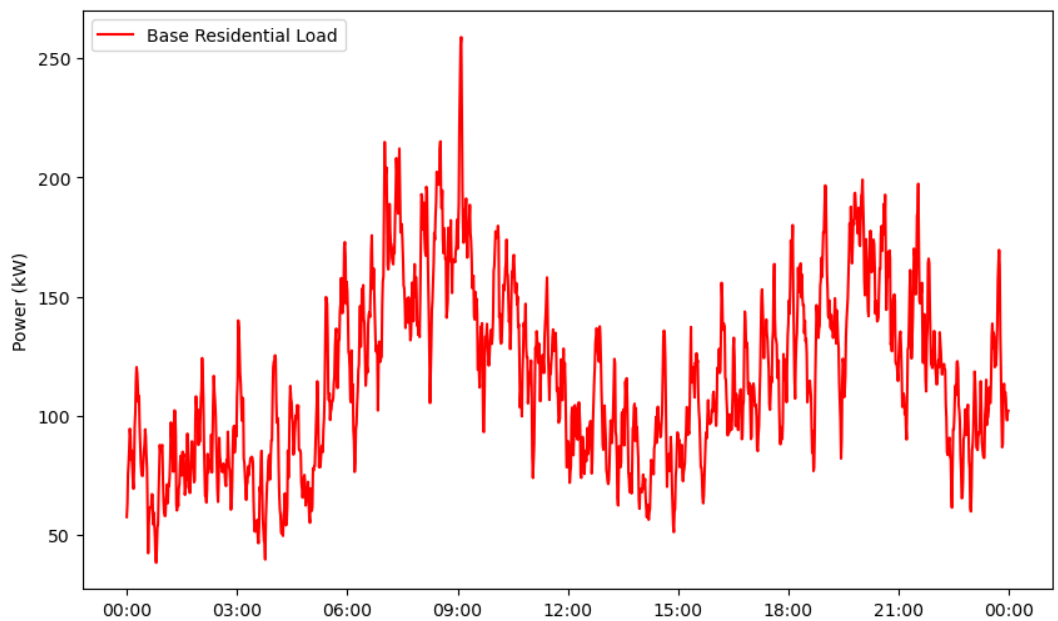

Figure 7-2: Aggregate demand for 114 single-family apartments on a weekday in July 2016 [99].

In addition to the charging session data, this study integrated time-dependent electricity prices obtained from the NPPD Utility. These prices were collected at various intervals throughout the day, enabling an analysis of the cost implications associated with EV charging. By taking into account the fluctuations in electricity prices, the study sought to evaluate the potential impact on the overall cost of charging EVs during different times of the day. Figure 7-3 illustrates the electricity price rates for a weekday in July, providing a visual representation of the varying prices throughout the day.

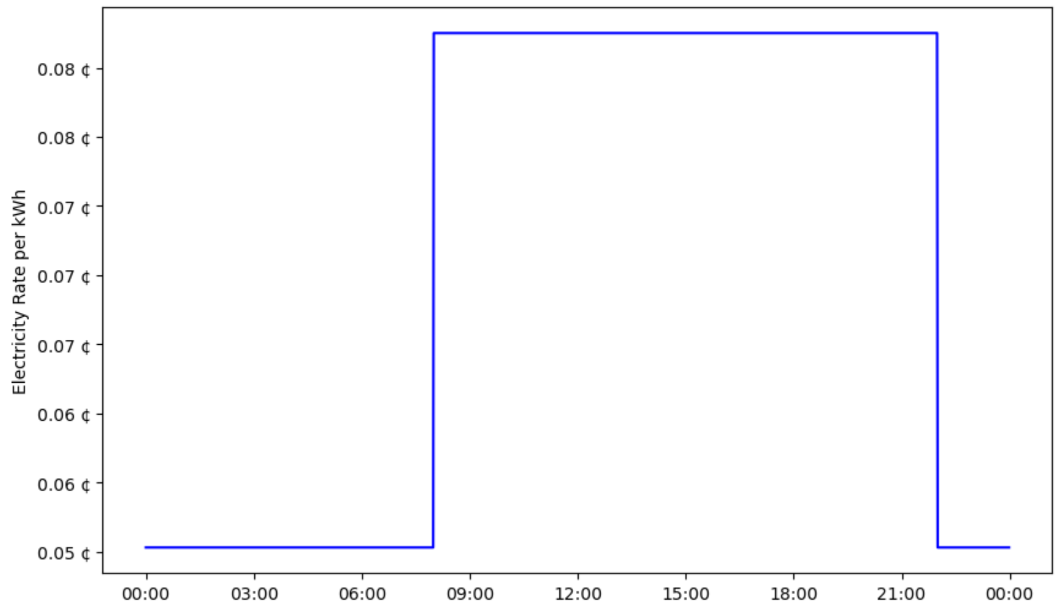

Figure 7-3: Electricity price rate for a weekday in July [100].

The charging session data was analyzed in detail, including start and end times, charging durations, and connection durations. Predictive values such as connection duration and charging duration were also calculated, which involved predicting connection end time and charging end time. To make these predictions, the Random Forest method was employed, using the most important features identified in Chapter 4. These features include mean, absolute time series, time of day, average frequency, time elapsed, and previous charging duration.

The data from the year 2021, which was used for the aggregate analysis in Chapter 5, was utilized for testing, while the remaining data was used for training and evaluation of the models. The evaluation metrics mentioned (MAE, RMSE, and R-squared) were calculated specifically for the testing data. They represent the performance of the trained models in predicting the target variables on unseen data from the testing dataset.

The prediction accuracy metrics for connection duration are as follows: The mean absolute error (MAE) is 246 minutes, the root mean square error (RMSE) is 335 minutes, and the R-squared value is 26%.

To optimize the charging schedule and manage the electricity demand, an optimization algorithm using CPLEX was employed. The optimization algorithm aimed to minimize the power capacity and the electricity cost for EV charging.

The optimization problem was formulated to allocate the charging durations of the EVs within the available time window, which was set to 1440 minutes representing a single day. If the predicted connection duration exceeded this time window, the connection duration was looped to the beginning of the day.

The CPLEX solver was used to solve the formulated optimization problem, considering constraints such as the available power capacity, EV charging requirements, and electricity prices at different time intervals throughout the day.

This mathematical model has been implemented in Python using DOcplex, the IBM Decision Optimization CPLEX Modeling for Python, and solved using CPLEX, a high-performance mathematical programming solver, version 20.1. The study was performed on an AMD A6-3400M APU @1.40 GHz. Because solving this problem using exact methods can be computationally intense, CPLEX was run until an optimality gap of less than 8.3% is achieved.

7.3 Results

The results section presents the findings of the EV charging optimization study. The objective of the optimization algorithm was to determine the optimal charging schedule for the EVs, considering factors such as peak demand and electricity price.

In the controlled charging scenario, where the optimization algorithm was applied, the total cost was observed to be 134$, while in the uncontrolled charging scenario (without optimization), the total cost was 147$. This indicates a difference of 13$ between the two scenarios. Moreover, the maximum power consumption during the charging period was 454 kW in the uncontrolled scenario, whereas it decreased to 361 kW in the controlled scenario as shown in Figure 7-4, resulting in a reduction of 93 kW in peak demand.

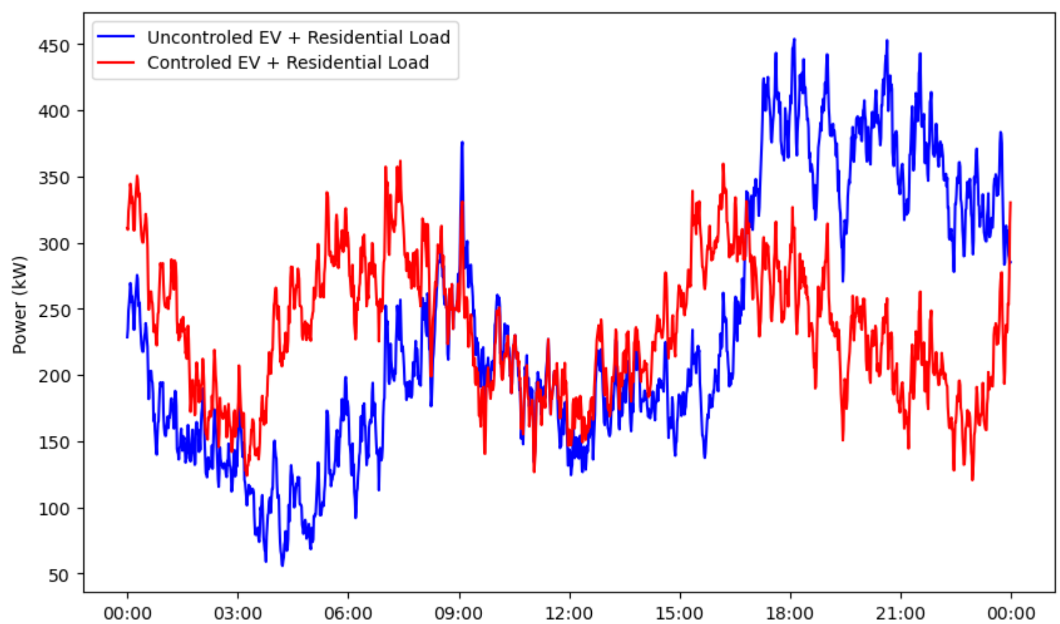

Figure 7-4: Uncontrolled load and controlled load for equal cost and peak demand weights

The comparison between the controlled and uncontrolled scenarios reveals the notable effectiveness of the optimization algorithm in reducing peak demand and minimizing electricity costs. With accurately predicted behavior, the implemented schedule achieves a remarkable 20.5% reduction in peak load and delivers substantial savings of 8.8% in electricity costs. However, actual charging and connection durations differ from the predictions due to various factors, such as unforeseen user behavior or external influences.

To evaluate the alignment between scheduled and actual charging sessions, the duration difference was calculated by subtracting the actual connection duration from the scheduled charging duration, accounting for looping. This analysis shed light on the discrepancies between the intended and actual charging periods for users.

Additionally, the absolute and percentage differences between the scheduled and actual charging durations were computed. Only 13.6% of users had fully satisfied sessions, indicating successful predictions and user adherence to the schedules. A significant portion of 77.8% of users experienced fully unsatisfied sessions, and 8.6% of users were only partially satisfied as shown in Figure 7-5. These results highlight the challenges of accurately predicting and accommodating individual charging behaviors.

Overall, the analysis reveals both strengths and limitations in the predictive models for EV charging behavior. While some users' charging patterns were accurately predicted, the majority experienced delays or dissatisfaction with the scheduled charging sessions. These discrepancies underscore the importance of continuous improvement in predictive models and the need for more personalized and user-centric charging strategies. The results provide valuable insights for refining the scheduling algorithms and enhancing the overall accuracy and efficiency of EV charging management. By addressing the observed discrepancies, the future of EV charging can be shaped to accommodate user preferences, optimize grid utilization, and promote sustainable and environmentally conscious transportation solutions.

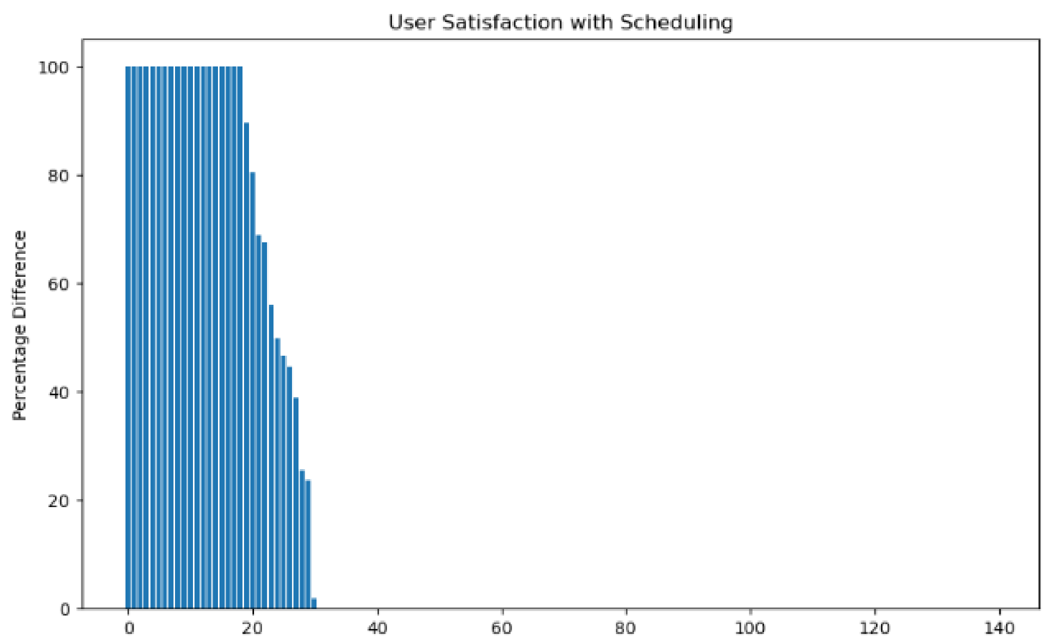

Figure 7-5: Percentage satisfaction of each charging session.

7.4 Conclusions

In conclusion, the evaluation of the scheduling optimization algorithm for EV charging sessions revealed the significant effect of prediction accuracy on the satisfaction of user demand. The algorithm aligned only approximately 13.6% of connections within the actual timeline. This finding indicates a need for further improvements in the scheduling process to minimize delays and enhance the overall user experience.

The assessment of alignment between scheduled and actual charging sessions through the duration difference analysis highlighted discrepancies between the intended and actual charging periods for users. The relatively low percentage 13.6% of fully satisfied sessions indicates the need for more accurate predictions and scheduling strategies tailored to individual charging behaviors.

These results underscore the importance of continuous refinement in predictive models and optimization algorithms to achieve a more user-centric and efficient EV charging management system. By addressing the challenges observed in this evaluation, such as reducing delays and enhancing predictive accuracy, the future of EV charging can be shaped to accommodate user preferences effectively and promote sustainable and environmentally conscious transportation solutions.

The evaluation of the scheduling optimization algorithm for EV charging sessions revealed both successes and challenges. While the algorithm demonstrated effectiveness in reducing peak demand and electricity costs, discrepancies between the predicted and actual connection periods lead to the majority of EVs not being fully charged. These differences emphasize the need for continuous refinement in predictive models and scheduling strategies to improve user satisfaction and ensure a reliable charging experience. Efforts to enhance predictive accuracy and user-centric scheduling will be crucial in shaping the future of EV charging and promoting sustainable transportation.

8. CONCLUSIONS AND FUTURE WORK

8.1 Conclusions

The objective of this research is to investigate the behavior of electric vehicle (EV) users in relation to charging and scheduling, develop an optimized EV demand scheduling algorithm, and explore the application of machine learning techniques for the prediction of EV charging demand. This study aimed to provide insights into user behavior, optimize scheduling algorithms, and showcase the benefits of machine learning in managing EV charging. The findings have important implications for the scheduling and management of EV charging infrastructure.

1. Understanding EV User Behavior

Through comprehensive analysis of EV user charging behavior in both household and public charging data, this research addresses a significant gap in the literature and emphasizes the impact of residential charging on the local electricity grid.

The analysis of EV charging session start times reveals distinct patterns in user behavior. In household charging, there is a gradual increase in sessions starting in the early morning, with a peak in the evening when users return home. Public charging stations show concentrated usage during the daytime, with prominent peaks at 8 am and 1 pm, likely corresponding to users arriving at their workplaces. Weekdays exhibit consistent charging routines, while weekends show variations in charging behavior.

The analysis of EV charging session end times provides insights into when users finish their charging sessions. In households, the majority of sessions end between 7-9 am, indicating completion before leaving for work. Public charging stations have a more even distribution of end times, with a peak at noon. Weekdays exhibit consistent morning disconnect times, while weekends display variability reflecting different charging behaviors and routines.

The analysis of connection durations in household charging data highlights that a significant portion (53%) of connections last 8 hours or more, with an average duration of 11.6 hours. This can be attributed to the dedicated usage of the charging station by the user, without any waiting time. In public charging data, 16.5% of connections also last 8 hours or more, with an average duration of 6.3 hours. Furthermore, the analysis of charging durations in both household and public charging sessions shows that the average duration for household charging is 2.4 hours, with only a small percentage (9.4%) of sessions exceeding five hours. Similarly, in public charging sessions, the majority (91.2%) have durations of five hours or less, with an average duration of 2.7 hours.

The analysis of energy consumption in the full dataset reveals interesting insights. It is observed that the majority (88%) of household charging sessions consumed less than 30 kWh, with an average energy consumption of 14 kWh. Similarly, 92% of public charging sessions consumed less than 30 kWh, with an average energy consumption of 12.7 kWh. These findings suggest that the majority of charging sessions exhibit relatively low energy consumption, with a significant portion falling below the 30-kWh mark. This information can be valuable for understanding the typical energy requirements of EV charging sessions and informing infrastructure planning and energy management strategies.

2. Optimized EV Demand Scheduling Algorithm

One of the key contributions of this research is providing valuable insights into EV charging scheduling and its implications. The quantitative findings highlight the benefits of strategic charging session management, presenting a solution for electric vehicle owners and utilities. By effectively balancing the objectives of minimizing electricity costs and reducing the impact on the electricity grid, scheduling optimization can create a more efficient and sustainable charging ecosystem. Another significant contribution of this study includes developing an optimization model for single-day EV charging scheduling, considering the trade-off between minimizing costs and reducing peak demand.

When 87% EV penetration is added to an existing residential load, the peak demand increases by 20% (91 kW) compared to the original load. The electricity cost rises by 22% ($88) for that day. Using an optimization model with equal weighting for cost and peak minimization, the peak demand decreases by 7% (516 kW) and there is a 13% ($11) cost saving in EV charging expenses. When only cost is minimized, the peak demand increases by 28% (146 kW) compared to the equal-weighting schedule, but there is still a 21% ($8) reduction in EV charging costs. However, when only the peak load is minimized, the peak demand decreases by 6% (489 kW) compared to the equal-weighting schedule. Unfortunately, the electricity cost increases by 15% ($12) compared to the equal-weighting schedule.

3. Application of Machine Learning Techniques

This study aimed to evaluate and compare the performance of different machine learning models in predicting various aspects of EV charging behavior. The Random Forest (RF) model emerged as the top performer in predicting charging demand, connection duration, and charging duration, achieving an R^2 value of 48% and lower error metrics compared to other models. However, accurately predicting these behaviors remains challenging, as indicated by the modest R^2 values and the need for further refinement.

The feature importance analysis highlighted key variables such as mean, average frequency, time of the day, time elapsed, previous charging value, and absolute time series that significantly influenced the predictions. Incorporating these variables improved the models' performance, but adding additional variables did not yield substantial improvements. The XGBoost algorithm showed relatively better performance in predicting the time until the next charge, but the results were still unsatisfactory, with an R^2 value of 21% and high error metrics.

Overall, accurately predicting EV charging behavior is complex due to its inherent variability. The results emphasize the relative superiority of the RF model, achieving an R^2 value of 48% and lower error metrics, and the importance of key variables in capturing patterns and improving prediction accuracy. Further research and refinement are needed to enhance the models' performance and develop more accurate predictive models for EV charging behavior.

4. EV User Behavior at Aggregated Level

Through comprehensive analysis of aggregated EV user behavior, This study shows the analysis of aggregated EV connectivity and power consumption data provides valuable insights into EV user behavior and charging patterns. The data reveals stable and consistent usage patterns throughout the year, with higher EV connectivity during winter months and fluctuations influenced by daily routines and holidays. Power consumption during EV charging shows similar patterns, with higher demand in winter months and fluctuations at monthly, weekly, and daily intervals.

Analyzing the data by averaging values for specific periods within a month, week, and day helps identify patterns and tendencies in EV connectivity and power consumption. This approach offers a comprehensive understanding of EV user behavior over time and aids in infrastructure planning to accommodate diverse charging needs.

The evaluation of the sessions data model for predicting aggregated charging behavior shows varying levels of accuracy at different time resolutions. The model performs reasonably well at longer time resolutions, such as weekly and monthly, with lower errors. However, at shorter time resolutions, such as minutes and hours, the model exhibits higher errors, indicating room for improvement.

5. Implications for EV Charging Infrastructure

The evaluation of the scheduling optimization algorithm for EV charging sessions revealed both successes and challenges. While the algorithm achieved alignment for approximately 13.6% of connections within the actual timeline, a notable number of connections experienced delays, with an average delay of 890 minutes. These findings highlight the need for further improvements in the scheduling process to minimize delays and enhance the overall user experience.

The assessment of alignment between scheduled and actual charging sessions through the duration difference analysis underscored discrepancies between the intended and actual charging periods for users. The relatively low percentage (13.6%) of fully satisfied sessions indicates the necessity for more accurate predictions and personalized scheduling strategies tailored to individual charging behaviors.

These results emphasize the importance of continuous refinement in predictive models and optimization algorithms to achieve a more user-centric and efficient EV charging management system. Addressing the observed challenges, such as reducing delays and enhancing predictive accuracy, will be critical in shaping the future of EV charging to effectively accommodate user preferences and promote sustainable and environmentally conscious transportation solutions.

8.2 Future Work

The comprehensive analysis conducted in this study has shed light on various aspects of EV charging behavior and scheduling optimization. Building upon the findings and insights gained from this research, several future research directions can be explored.

First, the development of more decentralized or personalized scheduling algorithms and charging strategies is essential to accommodate individual preferences for duration, timing, and location. By integrating user feedback and behavior analysis, these strategies can enhance user satisfaction and improve the overall charging experience. Additionally, considering user-specific constraints and requirements, such as preferred charging time windows and charging speed preferences, can further optimize charging schedules and better align them with user needs.

Another important direction for future research is the investigation of stochastic optimization approaches to EV charging scheduling. These approaches should not only consider the predicted charging behavior but also account for the inherent uncertainty present in the models. By incorporating probabilistic forecasts and scenario analysis, these approaches can provide more robust and adaptive charging schedules. Moreover, considering the uncertainty in factors such as electricity prices, renewable energy availability, and grid congestion can help optimize charging strategies and ensure optimal utilization of resources.

To improve session modeling, there is a need to tailor models to individual users and their specific charging behavior, despite computational challenges. By developing user-specific models, the unique patterns and characteristics of each user's charging sessions can be captured, leading to more accurate predictions and personalized charging strategies. This can involve incorporating additional user-specific factors, such as driving patterns, charging history, and preferences, into the modeling process. Advanced machine learning techniques, such as deep learning, can be explored to handle the complexity and variability of individual charging behaviors.

Furthermore, the analysis and prediction of user charging and connection times have broader implications for infrastructure planning and grid management. These insights can inform the modeling and optimization of Vehicle-to-Grid (V2G) systems, Micro-Grids, and the integration of renewable energy sources. Accurately predicting user behavior allows for the optimal planning and deployment of charging infrastructure, supporting renewable integration, grid stability, and efficient resource utilization. Additionally, understanding the impact of EV charging on grid congestion, peak demand, and load balancing can guide grid operators in managing EV charging to avoid infrastructure bottlenecks and optimize grid performance.

Finally, future research can explore the integration of demand response programs and dynamic pricing strategies into EV charging systems. By incentivizing users to adjust their charging behavior based on grid conditions, electricity prices, and renewable energy availability, demand response programs can effectively manage peak demand and promote grid stability. Dynamic pricing strategies aligned with grid dynamics can encourage off-peak charging and the utilization of renewable energy, ultimately reducing electricity costs for both EV owners and utility companies.

In conclusion, future work in the field of EV charging should focus on developing decentralized and personalized scheduling algorithms, leveraging stochastic optimization approaches, refining session modeling for individual users, exploring the broader implications of predicted user charging behavior on infrastructure planning, and integrating demand response and dynamic pricing strategies. These advancements will contribute to the efficient management of EV charging, improved user experience, grid stability, and the integration of EVs into a sustainable and resilient energy ecosystem.

9. REFERENCES

[1] B. G. Rabe, "Beyond Kyoto: Climate change policy in multilevel governance systems," *Governance*, vol. 20, no. 3, pp. 423–444, 2007, doi: 10.1111/j.1468-0491.2007.00365.x.

[2] R. S. Dimitrov, "The Paris agreement on climate change: Behind closed doors," *Glob. Environ. Polit.*, vol. 16, no. 3, pp. 1–11, 2016, doi: 10.1162/GLEP.

[3] C. Silva, M. Ross, and T. Farias, "Evaluation of energy consumption , emissions and cost of plug-in hybrid vehicles," *Energy Convers. Manag.*, vol. 50, no. 7, pp. 1635–1643, 2009, doi: 10.1016/j.enconman.2009.03.036.

[4] S. Shom, F. Al Juheshi, A. Rayyan, M. Alahmad, M. Abdul-hafez, and K. Shuaib, "Case Studies validating algorithm to determine the number of charging station placed in an Interstate and US-Highway," in *In 2017 IEEE International Conference on Electro Information Technology (EIT)*, 2017, pp. 50–55.

[5] "Global EV Outlook 2020." https://www.iea.org/reports/global-ev-outlook-2020.

[6] "Global EV sales and market share 2010 – 2021." https://www.greencarreports.com/news/1134999_global-ev-sales-more-than-doubled-in-2021-vs-2020-tripled-vs-2019.

[7] J. Sabata, S. Shom, A. Almaghrebi, A. McCollister, and M. Alahmad, "Incentivizing Electric Vehicle Adoption Through State and Federal Policies: Reviewing influential policies," *IEEE Electrif. Mag.*, vol. 11, no. 2, pp. 12–23, 2023, doi: 10.1109/mele.2023.3264889.

[8] A. Almaghrebi, X. Cheng, K. James, and M. Alahmad, "Analysis of PEV User Charging Behavior at Household Charging Stations , Omaha Case Study," in *2021 IEEE Transportation Electrification Conference & Expo (ITEC)*, 2021, pp. 561–566, doi: 10.1109/ITEC51675.2021.9490079.

[9] A. Bakhshinejad, A. Tavakoli, and M. M. Moghaddam, "Modeling and simultaneous management of electric vehicle penetration and demand response to improve distribution network performance," *Electr. Eng.*, vol. 103, no. 1, pp. 325–340, 2021, doi: 10.1007/s00202-020-01083-7.

[10] D. L. Boston, "Analysis of charging and driving behavior of plugin electric vehicles through telematics controller data," Georgia Institute of Technology, 2014.

[11] "Average Miles Driven Per Year: Why It Is Important." https://www.kbb.com/car-advice/average-miles-driven-per-year/.

[12] A. Almaghrebi, "The Impact of PEV User Charging Behavior in Building Public Charging Infrastructure," University of Nebraska-Lincoln, 2020.

[13] K. Mets, R. D'Hulst, and C. Develder, "Comparison of Intelligent charging algorithms for electric vehicles to reduce peak load and demand variability in a distribution grid," *J. Commun. Networks*, vol. 14, no. 6, pp. 672–681, 2012, doi: 10.1109/JCN.2012.00033.

[14] C. Michelbacher, "Grid Impacts of PEV Charging Infrastructure," 2018.

[15] M. Neaimeh *et al.*, "A probabilistic approach to combining smart meter and electric vehicle

charging data to investigate distribution network impacts q," *Appl. Energy*, vol. 157, pp. 688–698, 2015, doi: 10.1016/j.apenergy.2015.01.144.

[16] D. Mao, Z. Gao, and J. Wang, "An integrated algorithm for evaluating plug-in electric vehicle's impact on the state of power grid assets," *Electr. Power Energy Syst.*, vol. 105, pp. 793–802, 2019, doi: 10.1016/j.ijepes.2018.09.028.

[17] S. K. Hong, S. G. Lee, and M. Kim, "Assessment and mitigation of electric vehicle charging demand impact to transformer aging for an apartment complex," *Energies*, vol. 13, no. 10, p. 2571, 2020, doi: 10.3390/en13102571.

[18] J. J. Mies, J. R. Helmus, and R. van den Hoed, "Estimating the charging profile of individual charge sessions of Electric Vehicles in the Netherlands," *World Electr. Veh. J.*, vol. 9, no. 2, p. 17, 2018, doi: 10.3390/wevj9020017.

[19] S. Shao, M. Pipattanasomporn, and S. Rahman, "Challenges of PHEV penetration to the residential distribution network.," in *2009 IEEE Power & Energy Society General Meeting*, 2009, pp. 1–8, doi: 10.1109/PES.2009.5275806.

[20] H. Cai, X. Jia, A. S. F. Chiu, X. Hu, and M. Xu, "Siting public electric vehicle charging stations in Beijing using big-data informed travel patterns of the taxi fleet," *Transp. Res. Part D Transp. Environ.*, vol. 33, pp. 39–46, 2014, doi: 10.1016/j.trd.2014.09.003.

[21] T. Jonas, N. Daniels, and G. Macht, "Electric Vehicle User Behavior: An Analysis of Charging Station Utilization in Canada," *Energies*, vol. 16, no. 4, pp. 1–19, 2023, doi: 10.3390/en16041592.

[22] P. Jaruwatanachai, Y. Sukamongkol, and T. Samanchuen, "Predicting and Managing EV Charging Demand on Electrical Grids: A Simulation-Based Approach," *Energies*, vol. 16, no. 8, pp. 1–22, 2023, doi: 10.3390/en16083562.

[23] D. Kim, D. Kwon, J. Han, S. M. Lee, S. Elkosantini, and W. Suh, "Data-Driven Model for Identifying Factors Influencing Electric Vehicle Charging Demand: A Comparative Analysis of Early- and Maturity-Phases of Electric Vehicle Programs in Korea," *Appl. Sci.*, vol. 13, no. 6, p. 3760, 2023, doi: 10.3390/app13063760.

[24] A. Almaghrebi, S. Shom, F. Al Juheshi, K. James, and M. Alahmad, "Analysis of User Charging Behavior at Public Charging Stations," 2019, doi: 10.1109/ITEC.2019.8790534.

[25] A. Almaghrebi, F. Al Juheshi, J. Nekl, K. James, and M. Alahmad, "Analysis of Energy Consumption at Public Charging Stations, A Nebraska Case Study," in *2020 IEEE Transportation Electrification Conference & Expo*, 2020, pp. 1–6, doi: 10.1109/ITEC48692.2020.9161456.

[26] A. Almaghrebi, F. Aljuheshi, M. Rafaie, K. James, and M. Alahmad, "Data-Driven Charging Demand Prediction at Public Charging Stations Using Supervised Machine Learning Regression Methods," *Energies*, vol. 13, no. 16, p. 4231, Aug. 2020, doi: 10.3390/en13164231.

[27] X. Yang, Y. Cai, M. Zhang, and XiuYang, "Analysis of charging demand of household electric vehicles based on travel characteristics," in *2019 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia*, 2019, pp. 3217–3222.

[28] Z. Jiang, H. Tian, M. J. Beshir, R. Sibagatullin, and A. Mazloomzadeh, “Statistical Analysis of Electric Vehicles Charging , Station Usage and Impact on the Grid,” in *2016 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), Minneapolis, MN*, 2016, pp. 1–5, doi: 10.1109/ISGT.2016.7781228.

[29] C. Weiller, “Plug-in hybrid electric vehicle impacts on hourly electricity demand in the United States,” *Energy Policy*, vol. 39, no. 6, pp. 3766–3778, 2011, doi: 10.1016/j.enpol.2011.04.005.

[30] A. Ashtari, E. Bibeau, S. Shahidinejad, and T. Molinski, “PEV charging profile prediction and analysis based on vehicle usage data,” *IEEE Trans. Smart Grid*, vol. 3, no. 1, pp. 341–350, 2011, doi: 10.1109/TSG.2011.2162009.

[31] A. Foley, B. Tyther, P. Calnan, and B. Ó. Gallachóir, “Impacts of electric vehicle charging under electricity market operations,” *Appl. Energy*, vol. 101, pp. 93–102, 2013, doi: 10.1016/j.apenergy.2012.06.052.

[32] R. Wolbertus, M. Kroesen, R. van den Hoed, and C. Chorus, “Fully charged : An empirical study into the factors that in fluence connection times at EV-charging stations,” *Energy Policy*, vol. 123, pp. 1–7, 2018, doi: 10.1016/j.enpol.2018.08.030.

[33] S. Babrowski, H. Heinrichs, P. Jochem, and W. Fichtner, “Load shift potential of electric vehicles in Europe,” *J. Power Sources*, vol. 255, pp. 283–293, 2014, doi: 10.1016/j.jpowsour.2014.01.019.

[34] J. Schäuble, T. Kaschub, A. Ensslen, P. Jochem, and W. Fichtner, “Generating electric vehicle load profiles from empirical data of three EV fleets in Southwest Germany,” *J. Clean. Prod.*, vol. 150, pp. 253–266, 2017, doi: 10.1016/j.jclepro.2017.02.150.

[35] A. Lucas, R. Barranco, and N. Refa, “EV Idle Time Estimation on Charging Infrastructure, Comparing Supervised Machine Learning Regressions,” *Energies*, vol. 12, no. 2, p. 269, 2019, doi: 10.3390/en12020269.

[36] L. Buzna, P. De Falco, S. Khormali, D. Proto, and M. Straka, “Electric vehicle load forecasting: A comparison between time series and machine learning approaches,” *Synerg. MED 2019 - 1st Int. Conf. Energy Transit. Mediterr. Area*, pp. 1–5, 2019, doi: 10.1109/SyNERGY-MED.2019.8764110.

[37] S. Ai, A. Chakravorty, and C. Rong, “Household EV Charging Demand Prediction using Machine and Ensemble Learning,” in *2018 IEEE International Conference on Energy Internet (ICEI)*, 2018, pp. 163–168, doi: 10.1109/ICEI.2018.00037.

[38] Y. W. Chung, B. Khaki, T. Li, C. Chu, and R. Gadh, “Ensemble machine learning-based algorithm for electric vehicle user behavior prediction,” *Appl. Energy*, vol. 254, no. April, p. 113732, 2019, doi: 10.1016/j.apenergy.2019.113732.

[39] S. Shahriar, A. R. Al-Ali, A. H. Osman, S. Dhou, and M. Nijim, “Prediction of EV charging behavior using machine learning,” *IEEE Access*, vol. 9, pp. 111576–111586, 2021, doi: 10.1109/ACCESS.2021.3103119.

[40] A. Lilhore, K. K. Prasad, and V. Agarwal, “Machine Learning-based Electric Vehicle User Behavior Prediction,” *2023 IEEE IAS Glob. Conf. Renew. Energy Hydrog. Technol. GlobConHT 2023*, pp. 1–6, 2023, doi: 10.1109/GlobConHT56829.2023.10087780.

[41] S. Pokharel, P. Sah, and D. Ganta, “Improved prediction of total energy consumption and feature analysis in electric vehicles using machine learning and shapley additive explanations method,” *World Electr. Veh. J.*, vol. 12, no. 3, 2021, doi: 10.3390/wevj12030094.

[42] O. Frendo, N. Gaertner, and H. Stuckenschmidt, “Improving Smart Charging Prioritization by Predicting Electric Vehicle Departure Time,” *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 10, pp. 6646–6653, 2021, doi: 10.1109/TITS.2020.2988648.

[43] P. Date, “UC San Diego UC San Diego Electronic Theses and Dissertations Title Forecasting Electric Vehicle Arrival & Departure Time On UCSD Campus using Support Vector Machines,” 2017, [Online]. Available: https://escholarship.org/uc/item/1z43n3km.

[44] Y. Xiong, C. C. Chu, R. Gadh, and B. Wang, “Distributed optimal vehicle grid integration strategy with user behavior prediction,” *IEEE Power Energy Soc. Gen. Meet.*, vol. 2018-Janua, pp. 1–5, 2018, doi: 10.1109/PESGM.2017.8274327.

[45] Z. Fotouhi, M. R. Hashemi, H. Narimani, and I. S. Bayram, “A General Model for EV Drivers’ Charging Behavior,” *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 7368–7382, 2019, doi: 10.1109/tvt.2019.2923260.

[46] Q. Xing *et al.*, “Urban Electric Vehicle Fast-Charging Demand Forecasting Model Based on Data-Driven Approach and Human Decision-Making Behavior,” *Energies*, vol. 13, no. 6, p. 1412, 2020.

[47] M. Straka *et al.*, “Predicting popularity of electric vehicle charging infrastructure in urban context,” *IEEE Access*, vol. 8, pp. 11315–11327, 2020, doi: 10.1109/ACCESS.2020.2965621.

[48] J. Zhu, Z. Yang, Y. Chang, Y. Guo, K. Zhu, and J. Zhang, “A novel LSTM based deep learning approach for multi-time scale electric vehicles charging load prediction,” *2019 IEEE PES Innov. Smart Grid Technol. Asia, ISGT 2019*, pp. 3531–3536, 2019, doi: 10.1109/ISGT-Asia.2019.8881655.

[49] J. Zhu, Z. Yang, Y. Guo, J. Zhang, and H. Yang, “Short-term load forecasting for electric vehicle charging stations based on deep learning approaches,” *Appl. Sci.*, vol. 9, no. 9, 2019, doi: 10.3390/app9091723.

[50] S. Wang, L. Du, J. Ye, and D. Zhao, “A Deep Generative Model for Non-Intrusive Identification of EV Charging Profiles,” *IEEE Trans. Smart Grid*, vol. 11, no. 6, pp. 4916–4927, 2020, doi: 10.1109/TSG.2020.2998080.

[51] Y. Xiong, B. Wang, C. C. Chu, and R. Gadh, “Electric Vehicle Driver Clustering using Statistical Model and Machine Learning,” *IEEE Power Energy Soc. Gen. Meet.*, vol. 2018-Augus, pp. 1–5, 2018, doi: 10.1109/PESGM.2018.8586132.

[52] E. Xydas, C. Marmaras, L. M. Cipcigan, N. Jenkins, S. Carroll, and M. Barker, “A data-driven approach for characterising the charging demand of electric vehicles: A UK case study,” *Appl. Energy*, vol. 162, pp. 763–771, 2016, doi: 10.1016/j.apenergy.2015.10.151.

[53] A. Gerossier, R. Girard, and G. Kariniotakis, “Modeling and Forecasting Electric Vehicle Consumption Profiles,” *Energies*, vol. 12, no. 7, p. 1341, 2019, doi: 10.3390/en12071341.

[54] J. R. Helmus, M. H. Lees, and R. Van Den Hoed, “A data driven typology of electric vehicle user types and charging sessions,” *Transp. Res. Part C*, vol. 115, no. March, p. 102637, 2020, doi: 10.1016/j.trc.2020.102637.

[55] M. Giacomo, G. Prettico, A. Julea, G. Fulli, A. Mazza, and G. Chicco, “Statistical characterisation of the real transaction data gathered from electric vehicle charging stations,” *Electr. Power Syst. Res.*, vol. 166, no. September 2018, pp. 136–150, 2019, doi: 10.1016/j.epsr.2018.09.022.

[56] N. Sadeghianpourhamami, N. Refa, M. Strobbe, and C. Develder, “Electrical Power and Energy Systems Quantitive analysis of electric vehicle flexibility : A data-driven approach,” *Int. J. Electr. Power Energy Syst.*, vol. 95, pp. 451–462, 2018, doi: 10.1016/j.ijepes.2017.09.007.

[57] J. Quiros-Tortos, A. Navarro-Espinosa, L. F. Ochoa, and T. Butler, “Statistical representation of EV charging: Real data analysis and applications,” *20th Power Syst. Comput. Conf. PSCC 2018*, pp. 2013–2015, 2018, doi: 10.23919/PSCC.2018.8442988.

[58] B. Khaki, Y. W. Chung, C. Chu, and R. Gadh, “Nonparametric User Behavior Prediction for Distributed EV Charging Scheduling,” *IEEE Power Energy Soc. Gen. Meet.*, vol. 2018-Augus, pp. 1–5, 2018, doi: 10.1109/PESGM.2018.8585744.

[59] Z. Chen, Z. Zhang, J. Zhao, B. Wu, and X. Huang, “An Analysis of the Charging Characteristics of Electric Vehicles Based on Measured Data and Its Application,” *IEEE Access*, vol. 6, pp. 24475–24487, 2018, doi: 10.1109/ACCESS.2018.2835825.

[60] Y. W. Chung, B. Khaki, C. Chu, and R. Gadh, “Electric vehicle user behavior prediction using hybrid kernel density estimator,” *2018 Int. Conf. Probabilistic Methods Appl. to Power Syst. PMAPS 2018 - Proc.*, pp. 1–6, 2018, doi: 10.1109/PMAPS.2018.8440360.

[61] P. Thomas and P. K. Shanmugam, “A review on mathematical models of electric vehicle for energy management and grid integration studies,” *J. Energy Storage*, vol. 55, no. PA, p. 105468, 2022, doi: 10.1016/j.est.2022.105468.

[62] A. Almaghrebi, F. Vitor, K. James, F. Al Juheshi, and M. Alahmad, “A Framework for Scheduling Household Charging of Electric Vehicles,” 2022.

[63] S. Sabzi, L. Vajta, and T. Faghihi, “A Review on Electric Vehicles Charging Strategies concerning Actors Interests,” *Period. Polytech. Electr. Eng. Comput. Sci.*, vol. 66, no. 2, pp. 148–162, 2022, doi: 10.3311/PPee.19625.

[64] S. Z. Moghaddam and T. Akbari, “A Mixed-integer Linear Programming Model for the Plug-in Electric Vehicle Charging Problem in Unbalanced Low Voltage Electrical Distribution Systems Considering Neutral Conductor,” *Electr. Power Syst. Res.*, vol. 209, no. August 2021, p. 108049, 2022, doi: 10.1016/j.epsr.2022.108049.

[65] L. Hua, J. Wang, and C. Zhou, “Adaptive electric vehicle charging coordination on distribution network,” *IEEE Trans. Smart Grid*, vol. 5, no. 6, pp. 2666–2675, 2014, doi: 10.1109/TSG.2014.2336623.

[66] S. Xu, Z. Yan, D. Feng, and X. Zhao, “Electrical Power and Energy Systems Decentralized charging control strategy of the electric vehicle aggregator based on augmented Lagrangian method,” *Electr. Power Energy Syst.*, vol. 104, no. July 2018, pp. 673–679, 2019, doi:

10.1016/j.ijepes.2018.07.024.

[67] H. Hayajneh, M. B. Salim, S. Bashetty, and X. Zhang, "Logistics system planning for battery-powered electric vehicle charging station networks," *J. Phys. Conf. Ser.*, vol. 1311, no. 1, pp. 1–4, 2019, doi: 10.1088/1742-6596/1311/1/012025.

[68] Z. Miljanic, V. Radulovic, and B. Lutovac, "Efficient placement of electric vehicles charging stations using integer linear programming," *Adv. Electr. Comput. Eng.*, vol. 18, no. 2, pp. 11–16, 2018, doi: 10.4316/AECE.2018.02002.

[69] S. Rafique, M. S. H. Nizami, U. Bin Irshad, M. J. Hossain, and S. C. Mukhopadhyay, "EV Scheduling Framework for Peak Demand Management in LV Residential Networks," *IEEE Syst. J.*, pp. 1–9, 2021, doi: 10.1109/JSYST.2021.3068004.

[70] R. Das, Y. Wang, K. Busawon, G. Putrus, and M. Neaimeh, "Real-time multi-objective optimisation for electric vehicle charging management," *J. Clean. Prod.*, vol. 292, p. 126066, 2021, doi: 10.1016/j.jclepro.2021.126066.

[71] K. Rauma, A. Funke, T. Simolin, P. Järventausta, and C. Rehtanz, "Electric Vehicles as a Flexibility Provider: Optimal Charging Schedules to Improve the Quality of Charging Service," *Electricity*, vol. 2, pp. 225–243, 2021, doi: 10.3390/electricity2030014.

[72] P. Aliasghari, B. Mohammadi-Ivatloo, and M. Abapour, "Risk-based scheduling strategy for electric vehicle aggregator using hybrid Stochastic/IGDT approach," *J. Clean. Prod.*, vol. 248, p. 119270, 2020, doi: 10.1016/j.jclepro.2019.119270.

[73] J. Garcia-Guarin, W. Infante, J. Ma, D. Alvarez, and S. Rivera, "Optimal scheduling of smart microgrids considering electric vehicle battery swapping stations," *Int. J. Electr. Comput. Eng.*, vol. 10, no. 5, pp. 5093–5107, 2020, doi: 10.11591/IJECE.V10I5.PP5093-5107.

[74] F. Malandrino, C. Casetti, C. F. Chiasserini, and M. Reineri, "A game-theory analysis of charging stations selection by EV drivers," *Perform. Eval.*, vol. 83–84, pp. 16–31, 2015, doi: 10.1016/j.peva.2014.11.001.

[75] M. Zima-Bockarjova, A. Sauhats, L. Petrichenko, and R. Petrichenko, "Charging and Discharging Scheduling for Electrical Vehicles Using a Shapley-Value Approach," *Energies*, vol. 13, no. 5, p. 1160, 2020, doi: 10.3390/en13051160.

[76] C. P. Mediwaththe and D. B. Smith, "Game-theoretic electric vehicle charging management resilient to non-ideal user behavior," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 11, pp. 3486–3495, 2018, doi: 10.1109/TITS.2017.2784418.

[77] Z. Wang, D. Yue, J. Liu, and Z. Xu, "A Stackelberg Game Modelling Approach for Aggregator Pricing and Electric Vehicle Charging," *IEEE Int. Symp. Ind. Electron.*, vol. 2019-June, pp. 2209–2213, 2019, doi: 10.1109/ISIE.2019.8781294.

[78] M. Yoshihara and T. Namerikawa, "Non-cooperative Optimization of Charging Scheduling for Electric Vehicle via Stackelberg Game and Matching Theory," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 17029–17034, 2020, doi: 10.1016/j.ifacol.2020.12.1522.

[79] S. Hussain, M. A. Ahmed, K. B. Lee, and Y. C. Kim, "Fuzzy logic weight based charging scheme for optimal distribution of charging power among electric vehicles in a parking lot,"

Energies, vol. 13, no. 12, 2020, doi: 10.3390/en13123119.

[80] W. Jiang and Y. Zhen, "A Real-Time EV Charging Scheduling for Parking Lots With PV System and Energy Store System," *IEEE Access*, vol. 7, 2019.

[81] M. Elmehdi and M. Abdelilah, "Genetic algorithm for optimal charge scheduling of electric vehicle fleet," *ACM Int. Conf. Proceeding Ser.*, vol. Part F1481, no. March 2019, 2019, doi: 10.1145/3320326.3320329.

[82] V. Lazari and A. Chassiakos, "Multi-Objective Optimization of Electric Vehicle Charging Station Deployment Using Genetic Algorithms," *Appl. Sci.*, vol. 13, no. 8, 2023, doi: 10.3390/app13084867.

[83] G. Liu, L. Kang, Z. Luan, J. Qiu, and F. Zheng, "Charging station and power network planning for integrated electric vehicles (EVs)," *Energies*, vol. 12, no. 13, p. 2595, 2019, doi: 10.3390/en12132595.

[84] H. Ahmadi-Nezamabad, M. Zand, A. Alizadeh, M. Vosoogh, and S. Nojavan, "Multi-objective optimization based robust scheduling of electric vehicles aggregator," *Sustain. Cities Soc.*, vol. 47, no. February, p. 101494, 2019, doi: 10.1016/j.scs.2019.101494.

[85] K. Bakht, S. A. R. Kashif, M. S. Fakhar, I. A. Khan, and G. Abbas, "Accelerated Particle Swarm Optimization Algorithms Coupled with Analysis of Variance for Intelligent Charging of Plug-in Hybrid Electric Vehicles," *Energies*, vol. 16, no. 7, 2023, doi: 10.3390/en16073210.

[86] M. Alqahtani and M. Hu, "Dynamic energy scheduling and routing of multiple electric vehicles using deep reinforcement learning," *Energy*, vol. 244, p. 122626, 2022, doi: 10.1016/j.energy.2021.122626.

[87] J. Liu, H. Guo, J. Xiong, N. Kato, J. Zhang, and Y. Zhang, "Smart and Resilient EV Charging in SDN-Enhanced Vehicular Edge Computing Networks," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 1, pp. 217–228, 2020, doi: 10.1109/JSAC.2019.2951966.

[88] S. Morsalin, K. Mahmud, and G. Town, "Electric vehicle charge scheduling using an artificial neural network," *IEEE PES Innov. Smart Grid Technol. Conf. Eur.*, pp. 276–280, 2016, doi: 10.1109/ISGT-Asia.2016.7796398.

[89] T. Barker and A. Ghosh, "Neural Network-Based PV Powered Electric Vehicle Charging Station," *2022 IEEE Delhi Sect. Conf. DELCON 2022*, pp. 1–6, 2022, doi: 10.1109/DELCON54057.2022.9752821.

[90] Q. Dang, D. Wu, and B. Boulet, "A Q-Learning Based Charging Scheduling Scheme for Electric Vehicles," *ITEC 2019 - 2019 IEEE Transp. Electrif. Conf. Expo*, pp. 0–4, 2019, doi: 10.1109/ITEC.2019.8790603.

[91] J. Shi, Y. Gao, W. Wang, N. Yu, and P. A. Ioannou, "Operating Electric Vehicle Fleet for Ride-Hailing Services with Reinforcement Learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 11, pp. 4822–4834, 2020, doi: 10.1109/TITS.2019.2947408.

[92] B. Yagcitekin and M. Uzunoglu, "A double-layer smart charging strategy of electric vehicles taking routing and charge scheduling into account," *Appl. Energy*, vol. 167, pp. 407–419, 2016, doi: 10.1016/j.apenergy.2015.09.040.

[93] F. Sallabi, K. Shuaib, and M. Alahmad, "Online scheduling scheme for smart electric vehicle charging infrastructure," in *2017 13th International Wireless Communications and Mobile Computing Conference, IWCMC 2017*, 2017, pp. 1297–1302, doi: 10.1109/IWCMC.2017.7986472.

[94] R. Ghotge, Y. Snow, S. Farahani, Z. Lukszo, and A. van Wijk, "Optimized scheduling of EV charging in solar parking lots for local peak reduction under EV demand uncertainty," *Energies*, vol. 13, no. 5, p. 1275, 2020, doi: 10.3390/en13051275.

[95] S. S. Mohammed, T. P. I. Ahamed, S. H. E. A. Aleem, and A. I. Omar, "Interruptible charge scheduling of plug-in electric vehicle to minimize charging cost using heuristic algorithm," *Electr. Eng.*, 2021, doi: 10.1007/s00202-021-01398-z.

[96] T. Mao, X. Zhang, and B. Zhou, "Intelligent Energy Management Algorithms for EV-charging Scheduling with Consideration of Multiple EV Charging Modes," *Energies*, vol. 12, no. 2, p. 265, 2019.

[97] "Electric Vehicles Should Be a Win for American Workers." https://www.americanprogress.org/issues/economy/reports/2020/09/23/489894/electric-vehicles-win-american-workers/.

[98] A. Almaghrebi, F. Al, K. James, N. Aljuhaishi, and M. Alahmad, "PEVs Idle Time Prediction at Public Charging Stations Using Machine-Learning Methods," 2021.

[99] "Apartment dataset." https://traces.cs.umass.edu/index.php/Smart/Smart.

[100] "RateWise Time-Of-Use rate." https://www.nppd.com/rates/for-your-home.

[101] O. Kolawole and I. Al-Anbagi, "Electric vehicles battery wear cost optimization for frequency regulation support," *IEEE Access*, vol. 7, pp. 130388–130398, 2019, doi: 10.1109/ACCESS.2019.2930233.

[102] Z. Tian, L. Tu, C. Tian, Y. Wang, and F. Zhang, "Understanding Battery Degradation Phenomenon in Real-Life Electric Vehicle Use Based on Big Data," *Proc. - 2017 3rd Int. Conf. Big Data Comput. Commun. BigCom 2017*, pp. 334–339, 2017, doi: 10.1109/BIGCOM.2017.15.

[103] T. Chen and C. Guestrin, "XGBoost : A Scalable Tree Boosting System," in *In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[104] H. Fathabadi, "Plug-in hybrid electric vehicles: replacing internal combustion engine with clean and renewable energy based auxiliary power sources," *IEEE Trans. Power Electron.*, vol. 33, no. 11, pp. 9611–9618, 2018, doi: 10.1109/TPEL.2018.2797250.

[105] D. K. Choi, "Data-Driven Materials Modeling with XGBoost Algorithm and Statistical Inference Analysis for Prediction of Fatigue Strength of Steels," *Int. J. Precis. Eng. Manuf.*, vol. 20, no. 1, pp. 129–138, 2019, doi: 10.1007/s12541-019-00048-6.

[106] T. K. Ho, "Random decision forests," in *Proceedings of the 3rd International Conference on Document Analysis and Recognition*, 1995, pp. 278–282.

[107] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction.* 2009.

[108] "What are neural networks?" https://www.ibm.com/topics/neural-networks.

[109] A. D. Dongare, R. R. Kharde, and A. D. Kachare, "Introduction to Artificial Neural Network (ANN) Methods," *Int. J. Eng. Innov. Technol.*, vol. 2, no. 1, pp. 189–194, 2012, [Online]. Available: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1082.1323&rep=rep1&type=pdf.

[110] G. Asadollahfardi, "Artificial Neural Network," vol. 3, no. 1, pp. 77–91, 2015, doi: 10.1007/978-3-662-44725-3_5.

[111] A. T. C. Goh, "Back-propagation neural networks for modeling complex systems," *Artif. Intell. Eng.*, vol. 9, no. 3, pp. 143–151, 1995, doi: 10.1016/0954-1810(94)00011-S.

[112] D. K. Barupal and O. Fiehn, "Generating the blood exposome database using a comprehensive text mining and database fusion approach," *Environ. Health Perspect.*, vol. 127, no. 9, pp. 2825–2830, 2019, doi: 10.1289/EHP4713.

[113] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, vol. 13-17-August-2016, pp. 785–794, 2016, doi: 10.1145/2939672.2939785.

[114] T. K. Ho, "Random Decision Forests Tin Kam Ho Perceptron training," *Proc. 3rd Int. Conf. Doc. Anal. Recognit.*, vol. 1, pp. 278–282, 1995, [Online]. Available: https://ieeexplore.ieee.org/abstract/document/598994/.

[115] W. S. Mcculloch and W. Pitts, "A logical calculus nervous activity," *Bull. Math. Biol.*, vol. 52, no. l, pp. 99–115, 1990.