January 2023

Estimation Of Idle Time Using Machine Learning Models For Vehicle-To-Grid (V2G) Integration And Services

Prashanth Rajagopalan

Estimation of Idle Time using Machine Learning Models for Vehicle-to-Grid (V2G) Integration and Services

A Thesis Presented to The Academic Faculty by

Prashanth Rajagopalan

In Partial Fulfillment of the Requirements

For the Degree of

Masters of Science in the

School of Electrical Engineering and Computer Science

College of Engineering and Mines

University of North Dakota

Grand Forks, North Dakota

December 2023

Name: Prashanth Rajagopalan

Degree: Master of Science

This document, submitted in partial fulfillment of the requirements for the degree from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.

DocuSigned by:
Prakash Ranganathan
2E9D28B343204C6
Prakash Ranganathan

DocuSigned by:
Hossein Salehfar
83DF3018FE4944C
Hossein Salehfar

DocuSigned by:
Ediz Polat
6FF2EE17ED18483
Ediz Polat

This document is being submitted by the appointed advisory committee as having met all the requirements of the School of Graduate Studies at the University of North Dakota and is hereby approved.

DocuSigned by:
Chris Nelson
2E0AF088C793403...
Chris Nelson
Dean of the School of Graduate Studies

12/1/2023

Date

PERMISSION

Title:	Estimation of Idle Time using Machine Learning Models for Vehicle-to-Grid (V2G) Integration and Services
Department:	School of Electrical Engineering and Computer Science
Degree:	Master of Science

In presenting this thesis in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my thesis work or, in his absence, by the Chairperson of the department or the dean of the School of Graduate studies. It is understood that any copying or publication or other use of this thesis or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my thesis.

Name: Prashanth Rajagopalan

Date: 1^{st} December 2023

Contents

LIST OF FIGURES

LIST OF TABLES

Abbreviation

ACN	Adaptive Control Network
ADF	Augmented Dickey Fuller Test
AMI	Advanced Metering Infrastructure
ANL	Argonne National Laboratory
ANN	Artifical Neural Network
ANOVA	Analysis of Variance
AOA	Arithmetic Optimization Algorithm
AR	Auto Regressive
ARIMA	Auto Regressive Integrated Moving Average
BDL	Bayesian Deep Learning
CART	Classification and Regression Tree
CO_2	Carbon Dioxide
CNN	Convolutional Neural Network
CRPS	Continuous Ranked Probability Score
DBSCAN	Density Based Spatial Clustering of Application with Noise
DFNN	Deep Feedforward Neural Network
DKDE	Diffusion-based Kernel Density Estimator
DL	Deep Learning
DNN	Deep Neural Network
DoE	Department of Energy
DT	Decision Tree
ELM	Extreme Learning Machines

EMD	Emperical Mode Decomposition
EV	Electric Vehicles
EVSE	Electric Vehicle Supply Equipment
G2V	Grid-to-Vehicle
GAMS	Generic Algebraic Modeling System
GBM	Gradient Boosting Machines
GBR	Gradient Boosting Regressor
GMM	Gaussian Mixture Model
GRU	Gated Recurrent Unit
HEV	Hybrid Electric Vehicle
ICE	Internal Combustion Engine
IEA	International Energy Agency
IT	Idle Time
KNN	K-Nearest Neighbor
kWh	Kilowatt Hour
LR	Logistic Regression
LSTM	Long-Short Term Memory
MA	Moving Average
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MBQFN	Machine Based Quantile Forecast Network
ML	Machine Learning
MLP	Multi Layer Preceptron
MLR	Multiple Linear Regression
MToM	Machine Theory of Mind
MW	Megawatt
NAR	Non-linear Autoregressive Neural Network

NN	Neural Network
PCA	Principle Component Analysis
QR	Quantile Regression
RBF	Radial Based Function
RF	Random Forest
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
SARIMA	Seasonal Auto Regressive Integrated Moving Average
SARIMAX	Seasonal Auto Regressive Integrated Moving Average with Exogenous Factors
SoC	State of Charge
SSE	Sum of Squared Error
SVC	Support Vector Classifier
SVR	Support Vector Regression
ToU	Time of Use
V2G	Vehicle-to-Grid
V2H	Vehicle-to-Home
VTO	Vehicle Technologies Office
WHO	World Health Organization
XgB	XgBoost

ACKNOWLEDGEMENTS

Foremost, I want to express my gratitude to my advisor, Dr. Prakash Ranganathan, for his unwavering support and encouragement throughout my academic path. I appreciate the valuable insights I have acquired under his guidance.

Acknowledgments are also due to my thesis committee members, Dr. Hossein Salehfar, Professor, and Dr. Ediz Polat, Teaching Professor at SEECS, for their feedback and support during this research.

I extend my thanks to family, friends, and colleagues who stood by me during this journey. A heartfelt thanks goes to my parents, whose enduring support has consistently backed my ideas and interests.

ABSTRACT

Estimation of Idle Time using Machine Learning Models for Vehicle-to-Grid (V2G) Integration and Services

Prashanth Rajagopalan

As the Electric Vehicles (EVs) market continues to expand, ensuring the access to charging stations remains a significant concern. This work focuses on addressing multiple challenges related to EV charging behavior and Vehicle-to-Grid (V2G) services. Firstly, it focuses on accurate minute-ahead (20 minute & 30 minute intervals) load forecasts for an EV charging station by using four years of historical data, from 2018-2021. This data is recorded from a university campus garage charging station. Machine Learning (ML) models such as Seasonal Auto-Regressive Integrated Moving Average (SARIMA), Random Forest (RF), and Neural Networks (NN) are employed for load forecasts in terms of Kilowatt hour (kWh) delivered from 54 charging stations. Preliminary results indicate that RF method performed better compared to other ML approaches, achieving a average Mean Absolute Error (MAE) of 7.26 on historical weekdays data.

Secondly, it focuses on estimating the probability of aggregated available capacity of users for V2G connections, which could be sold back to the grid through V2G system. To achieve this, an Idle Time (IT) parameter was tracked from the time spent by the EV users at the charging station after being fully charged. ML classification methods such as Logistic Regression (LR) and Linear Support Vector Classifier (SVC) were employed to estimate the IT variable. The SVC model performed better in estimating IT variable with an accuracy of 85% over LR 81%.

This work also analyzes the aggregated excess kWh available from the charging stations for V2G services, which offer benefits to both EV owners through incentives and the grid by balancing the load. ML models, including Support Vector Regressor (SVR), Gradient Boosting Regressor (GBR), Long-Short Term Memory (LSTM), and Random Forest (RF), are employed. LSTM performs better for this prediction problem with a Mean Absolute Percentage Error (MAPE) of 3.12, and RF as second best with lowest 3.59, when considering historical data on weekdays.

Furthermore, this work estimated the number of users available for V2G services corresponding to 15% and 30% of excess kWh, by using ML classification models such as Decision Tree (DT) and K Nearest Neighbor (KNN). Among these models, DT performed better, with highest 89% and 84% accuracy respectively.

This work also investigated the impact of the COVID-19 pandemic on EV users' charging behavior. This study analyzes the behavior modelled as before, after, and during COVID-19, employing data visualization using K-means and hierarchical clustering methods to identify common charging pattern with connection and disconnection time of the vehicles. K-means clustering proves to be more effective in all three scenarios modeled with a high silhouette index. Furthermore, prediction of collective charging session duration is achieved using ML Models, RF and XgBoost which achieved a MAPE of 14.6% and 15.1% respectively.

Keywords: Electric Vehicles, Vehicle-to-Grid (V2G), Idle time, Machine Learning.

Chapter 1

Introduction and State-of-the-art Methods for V2G Services

1.1 Introduction

In today's world, Electric Vehicles (EVs) are promising towards objective of sustainability. The popularity of EV among consumers are steadily rising, attributed with its extended battery range. According to the report from International Energy Agency (IEA), EV sales reached a milestone in 2021, doubling to 6.6 million units compared to its previous year [1]. Additionally, an exclusive analysis by Alliance for Automotive Innovation Report reveals that in Q1 2023, more than 305,000 units of EVs were sold in the United States, marking an increase of 56% compared to the same period in 2022. The top five states in the U.S. for EV sales in Q1 2023, includes California, leading with 23.9% followed by District of Columbia (20.1%); Washington (16.9%); Oregon (16.0%) and Nevada (14.9%) [2]. According to the report from Argonne National Laboratory (ANL), the sales of Hybrid Electric Vehicles (HEVs) have surged up to 75% in the United States, from comparing the sales in July 2022 [3].

As the EV market expands, the utilization of public charging stations is on the rise. According to the U.S. Department of Energy (DOE) and Vehicle Technologies Office (VTO), the United States is working towards a goal of setting up a network of 1.2 million public chargers by 2030. For this goal to be achieved, the Biden-Harris Administration have committed nearly $24 billion [4]. Figure 1.1. illustrates the growth of EV charging station location and EVSE from the period of 2017-2022. In the first quarter of 2023, the number of Electric Vehicle Supply

Equipment (EVSE) ports in the station locator increased by 3.2%, as equivalent to 5,047 EVSE ports. In particular, the South-Central region had 7.9% increase in public charging infrastructure in Q1, while California, which accounts for nearly one-third of public charging infrastructure of the country, continues to lead the country in the number of public ports. The number of Level 2, DC Fast public EV chargers increased 31% year-over-year, from 101,946 in the first quarter of 2022 to 133,925 in first quarter of 2023 [5]. Figure 1.2. represents the number of Level 2 and DC fast charge port count from 2017-2022 across the U.S.

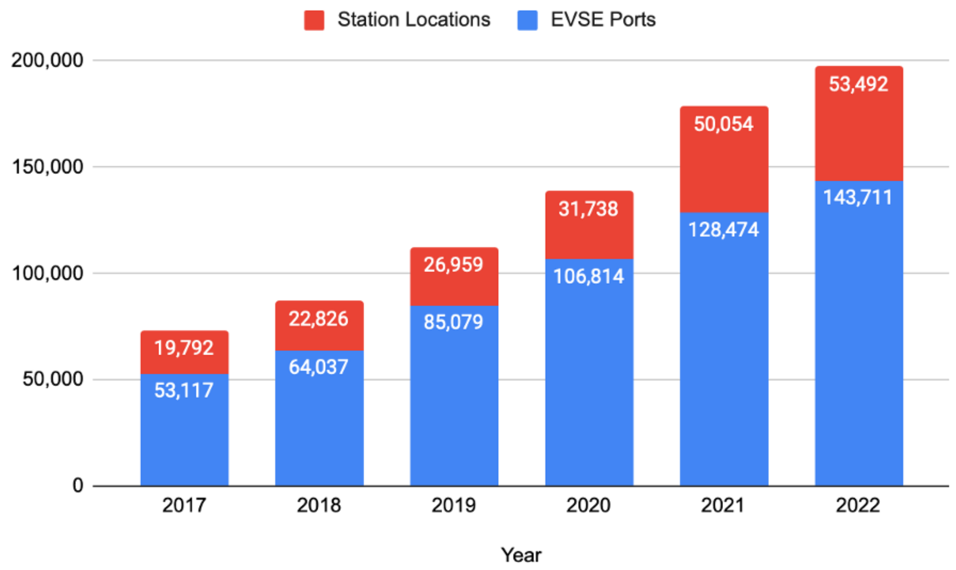


Figure 1.1: U.S. Public and Private Electric Vehicle Charging Infrastructure Growth from 2017-2022.

EVs with the qualification of newly purchased in or after 2023, with battery capacity of at least 7 kWh, will be eligible for a clean vehicle tax credit up to $7,500 under Internal Revenue Code Section 30D. Additionally, used vehicles purchased in 2023 or after are eligible for a tax credit of up to $4,000 [6].

EVs are widely considered to be environmentally friendly when compared to gasoline-powered vehicles, particularly in terms of carbon-dioxide (CO_2) emissions that contributes to climate change. According to the Congressional budget office, CO_2 emission is the largest source of greenhouse gases in the U.S. transportation sector [7]. In 2021, CO_2 emissions in the transportation sector has

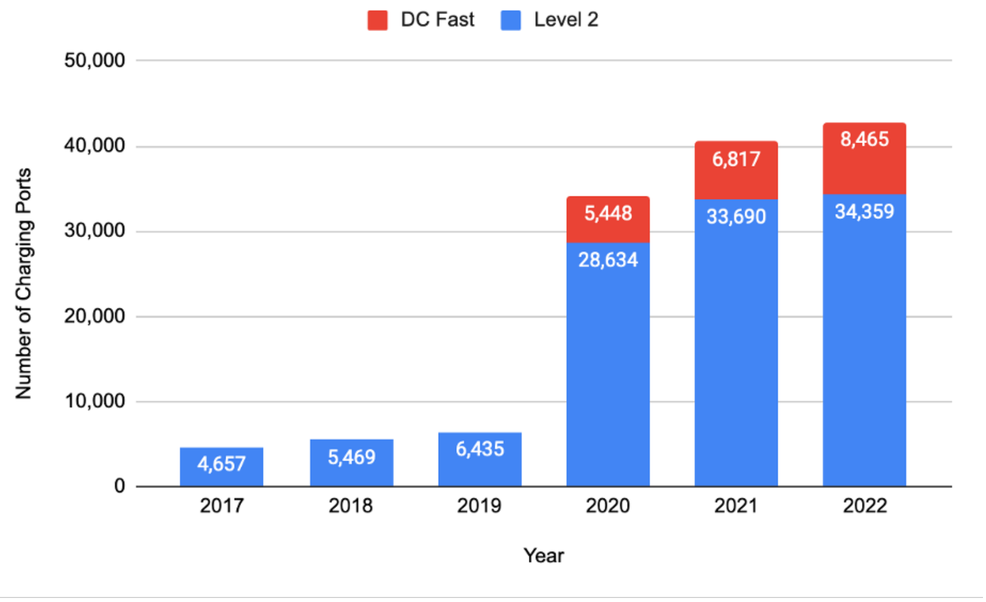


Figure 1.2: Number of Level 2 and DC Fast Charge Port from 2017-2022 across the U.S.

decreased 6% less compared to the scenario in 2005. Motor vehicles are solely responsible for 83% of CO_2 emissions in the transportation sector during 2019. Emissions from transportation sector surpassed those from the electric power sector five years ago, and it now constitutes 40% of total domestic emissions from the combustion of fossil fuels [8].

According to the public power association, the public utilities commission of the State of California have released a report on V2G integration working group proposing strategies in improving grid resiliency and security [9]. Widely deployed, V2G technology have the potential to aggregate EVs to act as a virtual power plant and energy storage system. This technology can effectively help in public safety during power shutoff events [10]. EVs and the charging stations are important assets in fighting the climate change. Therefore, it necessitates the study of charging station operation consisting of incorporating parameters like charger utilization and energy delivered.

1.2 Related Works

Investigating EV charging behavior pattern, load forecasts in charging stations, and idle time estimation are all important aspects in managing EV charging infrastructure. In recent years, data-driven approaches that use ML are becoming popular to solve charging infrastructure planning problems as well as the utilization of charging. The section highlights the state-of-the-art methods used in the existing research works in studying the charging behavior pattern, load forecasting, and grid integration parameters.

1.2.1 Study of EV User Charging Behavior

Sun et al. used DL approaches to forecast EV charging behavior [11]. The authors have analyzed the datasets by using K-means clustering method. The prediction results shows the stability of charging behavior divided into regular and irregular users for a charging station from Los Angeles. Based on this analysis, the authors have proposed LSTM based EV charging behavior prediction model and achieved prediction accuracy of 98.41%. In [12], the authors have utilized the user's charging pattern attributes such as charging cost, station accessibility to analyze the market penetration of EV. Additionally, the authors have considered consumer preferences, charging time of day, and types of EVSE preferred. They presented demand impact forecasts using a regression-based consumer utility function. In [13], the authors have conducted a survey in Germany, presenting a two-stage (qualitative survey in stage 1, followed by simulation in stage 2). This study addresses the impact of user characteristics such as socio-economic and demographic aspects to compare the charging behavior by EV and fuel refilling by ICE users. The results indicate that the users' decision varies with the factors such as range availability and financial condition for users, while choosing EV over ICE vehicles. Within the EV fleet, the effectual charging cycle scenarios needs to be weighed with its optimal V2G control strategies. Tang et al. proposed an optimal EV charging and V2G control strategy algorithm based on aggregated demand levels [14]. The authors have considered the State of Charge (SoC), charging rate per hour and departure time, this EV modelling and V2G scenario achieved significant cost reduction in the total charging cost of the EVs. Wu et al. proposed an optimal charging strategy with respect to the battery SoC, to predict the day-ahead energy market [15]. The findings indicate the EV aggregators are capable of providing up/down power reserves with the regular charging cycles from the EV owners. Singh et al. evaluated the behavioral factors such as risk-awareness and knowledge gap for slow EV adoption rate [16]. The authors have used convex optimization algorithm on a microgrid level scenario with active and inactive users to study the energy characteristics, economic criteria and battery characteristics in V2G and Vehicle-to-Home (V2H) scenarios.

The impact of the pandemic caused a significant shift in EV charging behavior. Wen et al. created an evidential survey on chinese EV industry [17]. Based on the important factors and features such as global EV sales, COVID-19 cases, active daily users, the findings indicate that the pandemic have impacted for the short-term as it disrupted the production at large scale during the year 2020 and early 2021. Shahriar et al. utilized K-means, hierarchical, and Gaussian Mixture Models (GMM) to find clusters of charging behavior during the pandemic [18]. The authors have used Silhouette, Davis-Bouldin and Calinski-Harabasz as internal validation indices and have achieved 0.41 as the best in silhouette indices. To study the charging session type and user behavior, Helmus et al. have found 13 distinct charging session types from a real-time charging station [19]. The authors have used GMM model and partitioned it to clusters of daytime and overnight charging sessions based on connection time and duration between two sessions, to study the behavior of users at the charging station. The results indicate 96.5% of total regular users of charging station. Khan et al. compared the performance of ML and DL approaches on the ACN dataset to predict charging behavior [20]. The results prove that KNN model outperformed NN in session duration prediction and RF performed better for energy demand prediction. To understand the charging behavior of EV users, Patrick et al. conducted statistical analysis such as Analysis of Variance (ANOVA) and other statistical tests to evaluate on charge event data recorded by data loggers located at the charge points throughout Ireland [21]. Based on the charge consumption and its duration, the authors reveal that users prefer to charge the EVs at home during the evening, which conflicts with the peak grid demand period. Similarly, Viswanathan et al. developed an assessment model to predict the utilization of public charging infrastructure throughout the City of San Diego [22]. Based on the travelling distance, battery capacity, the authors have formulated the scenarios of energy distribution and have achieved R-square value of 0.25 in the analysis. To assess the impact of the pandemic, Kanda et al. conducted a study on the economic effects of COVID-19 and how significantly it reduced EV sales in the short-term [23]. They also studied the opportunities in sustainable transition in the EV market of Finland and Sweden. This study shows the importance of adaption of EVs amidst the pandemic.

1.2.2 Estimation of Idle Time

To predict the EV user behavior for real-time charging data with 252 users, ML models, RF, SVR, and Diffusion-based Kernel Density Estimator (DKDE) models were employed [24]. The data is based on idle time and energy consumption for the UCLA campus charging station, containing connect and disconnect time data from 2015-2017. The authors have highlighted that SVR yields better MAPE over other methods for estimating Idle time. Flammini et al. have utilized a Beta Mixture Model approach on a real-time dataset from Netherlands containing 400,000 EV charging transactions containing driving behaviors patterns, demand, idle time, power and energy [25]. The findings indicate that 50% of the charging cycle last for less than 4 hours; the idle time depends on the geographical location and last less than four hours. Sadeghianpourhamami et al. studied user's arrival and departure times using three behavioral clusters using Density Based Spatial

Clustering of Application with Noise (DBSCAN) method for different scenarios such as charging at home and public charging in Netherlands [26]. The results highlight the mean idle time in each scenario modelled as park to charge with aggregated 101 hours, charge at home with 109 hours and charge at work with 115 hours of aggregated idle time. Similarly, the length of a charging session at public charging stations was studied by Rick et al. and their results show that the connection time exceeds the time needed to recharge the EV's batteries. Hence the charging session duration prediction plays a crucial role in determining the duration information [27]. As this study suggest, it is not only reliant on the charging duration but also on the occupancy of a particular parking lot for a longer duration which may affect the other incoming users' convenience. Figure 1.3. represents the various ML and DL approaches utilized in the background study.

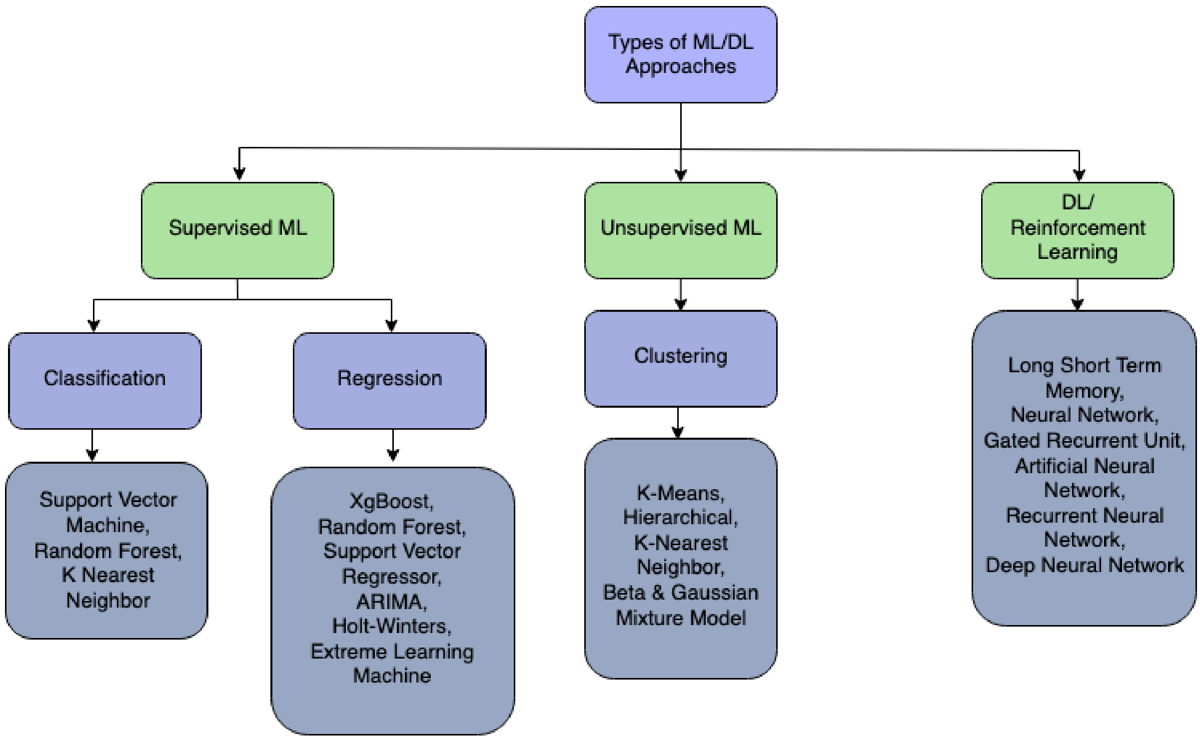


Figure 1.3: Types of ML & DL Approaches Utilized in the Literature Review.

1.2.3 Load Forecasting and V2G in Demand Response Management

Hu et al. proposed Machine Theory of Mind (MToM)- Based Quantile Forecast Network (MBQFN) on the ACN dataset to estimate the EV charging demand. To evaluate the performance, a Continuous Ranked Probability Score (CRPS) metric is utilized. The focus of this study is to forecast the quantiles of future charging demand in a charging station into 5 minutes ahead forecasts and have achieved 82.2% [28]. Power consumption is one of the important parameters to consider in the analysis of load characteristics. The demand forecast for the charging stations, on a broader scale are as important as determining the load characteristics from a residential setup. Liu et al. produced a 15-minute demand forecasts using a hybrid model containing Holt-Winters and Extreme Learning Machine (ELM) methods for residential electricity consumption data based on its seasonality. The results based on these models for 10-day and 30-day training set produces better accuracy than LSTM, with MAE 13.65 and 36.27 [29]. Almaghrebi et al. used a dataset containing 7 years of charging session, recorded from a public charging station in Nebraska. The authors employed ML methods such as XgB, SVR, RF to predict the EV charging demand. The parameters of interest from the dataset includes historical weekdays and energy utilized from the charging station. This study shows that RF and XgB have MAE of 4.07 and 4.12 respectively [30]. To predict the energy consumption and financial characteristics for effective V2G strategy, Connor et al. considered variables such as EV model, peak, off-peak time, battery life, supply/demand intensities under different scenarios of EV dataset from a university campus located in UK [31]. The findings show that utilizing energy stored from the EV batteries costs 64.7% less and also it costs 9.79% lesser than purchasing from the grid. Li et al. modelled a market distribution mechanism to study the net profit from V2G scenario. The results show that EV users possess cost-benefit by participating in V2G peak shaving services for users with low-cost batteries. These users are more likely to participate in V2G activities [32]. In [33], authors used Swedish driving and energy consumption patterns to develop a V2G econometric model, for multiple regions in Europe. The findings reveal that by properly integrating V2G services, Europe can cut 50% in investment costs during peak hours and EVs could act as a source in flexible demand response management. Zhou et al. considered the temporal and spatial characteristics of daily EV charging loads and predicted the similar daily loads using hybrid CNN-LSTM model. Before feeding into the CNN-LSTM, the authors used combination of Principal Component Analysis (PCA) and density peak clustering scheme to extract similar loads from raw charging data [34]. Malley et al. proposed a charge scheduling framework from aggregated bidirectional chargers with respect to the connect and disconnect times. This forecasting approach have utilized the charging events based on EV connection time series in to 5-minutes period and separated them into weekdays and weekends to achieve the probabilistic forecast for EV connection and frequency response during an outage [35]. Nogay et al. developed a LSTM and NN based prediction model from aggregated available capacity from 7 EVs for a ten-day travel created with specific driving records to demonstrate the market events. The results with V2G contribution from the simulation indicates the contribution of 1.2708 kWh with 33 half-hours time steps. Hence the concept of aggregated available capacity of users for V2G services plays a major role [36].

In order to investigate the feasibility of V2G operations, Gioradano et al. proposed an automated aggregator algorithm with day-ahead optimization of an EV charge fleet [37]. Based on clustering analysis of 215 EV users' behavior from 2013-2015, the results indicate a potential cost reduction up to 57%. Additionally, the findings suggest that it is feasible to utilize 50% of the aggregated battery capacity for V2G services. Similarly, Gautam et al. used Multi-layer Perceptron (MLP) to investigate the Time-of-Use (ToU) and cost function for V2G services. The MLP model utilized EV State of Charge to reduce peak demand and valley filling [38]. This hourly load forecasting problem incorporated with meteorological factors deployed MLP approach to achieve a MAPE of 4.11%. Mohanty et al. used Support Vector Regressor (SVR) on Advanced Metering Infrastructure (AMI) data for charge scheduling problem [39]. In this study, the authors predict home charge scheduling for V2G and Grid-to-Vehicle (G2V) scenario based on EV SoC. The results achieved a test set accuracy of 94.86% for the V2G case. In [40], simulated EV charging pattern data with features such as period of demand, supply and energy pricing were used to optimize the EV charge schedule. The results indicate 24% reduction in annual EV demand. Rezaei et al. proposed a load forecasting model to examine intelligent management of EV charging. This work utilized Generic Algebraic Modeling System (GAMS) for simulating 65 available parked EVs for analyzing the charging patterns [41]. This method of EV charging and discharging technique is designed for a commercial building in Detroit, and their results include a 16.6% reduction in the peak demand curve. Cui et al. proposed a DL-based demand forecasting model to study the factors affecting user behavior pattern [42]. The authors used MATLAB simulation on daily load curve regarding EV charging for a 10-month period. The authors have compared the performance of DL model with ARIMA, and the DL model achieved a prediction accuracy which is 5%-10% better compared to ARIMA model.

In [43], a day-ahead load forecasting technique using Deep Feedforward Neural Network (DFNN) and an Extreme Learning Machine (ELM) models were proposed. The result have achieved a lowest MAPE score of 4.16% for 5-minutes ahead forecasts, for the EV load data recorded for 11 days during the year 2019. Similarly, Van Kriekinge et al. used an LSTM-based algorithm to analyze day-ahead EV charge load with a 15-minutes ahead interval [44]. This work compared the EV charging behavior depending on the dynamics of the calendar and weather features. Given these characteristics, the authors assert that the RMSE was reduced to 16.16% from 19.22%. Shang et al. proposed a framework based on Deep Learning (DL) and ML models in predicting the V2G capacity from an EV fleet [45]. The authors assert that using LSTM with a cloud-edge based framework decreases the computational complexity in estimating the V2G dispatch. Considering 15-minute time steps, the prediction results for the LSTM model achieved 95% accuracy. Boulakhbar et al. conducted a comparative analysis of four established DL models: LSTM, Recurrent Neural Networks (RNN), Artificial Neural Networks (ANN), and Gated Recurrent Units (GRU). The authors performed the study for 2000 charging sessions recorded at a charging station in Morocco [46]. The results indicate that GRU performed better in predicting load demand, with an 2.90 RMSE and 0.76 MAPE for the test data. Similarly, Ma et al. used LSTM model for analyzing the occupancy of EV charging stations in Dundee, UK with three different types of charging, slow, fast and rapid charging [47]. The prediction accuracy is computed for 10 minutes, and hour ahead intervals. The LSTM model performed better with 81.87% accuracy for a 10-minute ahead forecast.

The analysis of short-term demand forecasting plays a crucial role in estimating the amount of V2G contribution from the EV charging networks. Zhu et al. used DL-based approach, DNN, RNN, LSTM and GRU, for hourly short-term load forecasting [48]. The authors conducted an analysis of charging load data from April 2017 to June 2018. The preliminary results provided a MAE score of 1.32, 0.91, 0.90, 0.70 for the respective models. Lu et al. proposed a Classification and Regression Tree-based (CART) RF algorithm to predict the short-term demand from the charging station [49]. The authors validated the results by using a practical charging system scenario over 6000 charging piles built in Shenzhen. The results achieved a MAPE of 9.76% and an RMSE of 2.27. Sun et al. deployed the SVR approach to forecast the charging load from JZ EV charging station dataset from Shandong [50]. The historical load data includes charging sessions, EV SoC on weekdays, weekends, and meteorological variables. The authors has achieved an MAE of 103.868 and an RMSE value of 149.86. Wang et al. proposed a charge demand forecasting approach with charging time probability estimated by Monte Carlo simulation [51]. This study was applied on charging station from Shenzhen city. The prediction of charging load highlights three-peak patterns based on vehicles that reached at 9AM, 3PM and 8PM having 15 Megawatt (MW), 16 MW and 22 MW respectively. The results suggest that power grids can accommodate low-cost charge scenarios in the early hours of the day.

To forecast the aggregated load from the charging station, Gerossier et al. have utilized data from the individual meters at charging stations and identified four main groups of charging behaviors using hierarchical clustering [52]. The clusters identified from the user groups indicates that there is frequent charging up to 52% occupancy during the night time, followed by evening/late evening. For implementing a smart EV charging algorithm to perform charge scheduling with respect to users' price preferences, Wang et al. predicted the day-ahead charging cost and energy consumption using ARIMA model for the charging station data from UCLA [53]. The cost and energy consumption comparison show that users with longer idle time possess better bidding strategy and can save the charging cost. Youssef et al. studied the distribution of charging session using heat maps based on the connection duration of each session for a public charging station in the Netherlands. A clustering technique was used to assess the smart charging potential [54]. The authors claim that 40% of the aggregated sessions occurs in the evening between 4PM to 8PM which is beneficial for the users in avoiding obstacles from peak hours. These studies provides the significance of load forecasting, idle duration estimation and excess energy prediction, additionally, recent literature review on load forecasting, charging demand, SoC estimation are illustrated in Table 1.1. The important findings and comparison with this work are highlighted in the table.

Table 1.1: Literature Review of Recent Research and Analysis on EV Charging Systems.

Dataset	**Location**	**ML Models**	**Metrics**	**Reference & Year of Publication**
3395 EV charging sessions, 85 EV drivers with repeat usage at 105 stations.	Across 25 sites at a workplace charging program.	DT, RF, SVC, **KNN**, DNN, LSTM	Accuracy - 93%, 94%, 29%, **41%**, 77%, 94%.	[55], 2023
NTS data containing travel patterns in the UK from 2002 to 2019.	United Kingdom	Light GBM	Accuracy - 85.8%	[56], 2022
Number of EVs - 1500, 2000 and 2500 with battery capacities in the range of 15-19 kWh to predict State of Charge	India	ANN, XGB	MAPE - 1.45, 2.14	[57], 2022
72 real driving trips with a BMW from Battery and Heating Data in Real Driving Cycles	Germany	**LSTM** and Nonlinear Autoregressive Neural Network (NAR)	RMSE, MAPE, MAE: LSTM - 0.7167, 0.0404; NAR - 0.5487, 0.0255, 1.8452 2.7665	[36], 2022
EV charging data from 25 public charging stations	Boulder, Colorado	**LSTM**, RNN, **SARIMA** and ARIMA,	RMSE, MSE: LSTM - 0.036, 0.026; RNN - 0.367, 0.229; SARIMA - 1.06, 0.819 ARIMA - 0.831, 0.643	[58], 2023
EV Trajectory dataset with 1000 charging stations with 76,000 private EVs from January 2018	Beijing, China	ARIMA, MLP, **LSTM**	MAPE: ARIMA - 21.57, MLP - 18.31, **LSTM - 6.83**	[59], 2023
EV Charging data collected by a charging station on Georgia Tech campus	Georgia, USA	Emperical Mode Decomposition (EMD) - Arithmetic Optimization Algorithm (AOA), LSTM	MAE - 0.1083	[60], 2022
Charging session data of multiple EV charging stations in a 35 kV power supply area	China	LSTM, SVM, ARIMA, SARIMAX	MAE: **LSTM** - 1996.23, 2360.43, 2480.47, **2171.76**	[61], 2022
Charging station with 64 parking spaces for electrified buses, 12 charging spaces for cars from March 2017	Shenzhen, China	ANN, RNN, LSTM, Enhance attention based **LSTM**	MAE: ANN - 267.904, RNN - 244.101, LSTM - 218.393, EA-LSTM - **159.659**	[62], 2022
EV charging session data collected on the Caltech campus	Psaedena, California	Multiple Linear Regression (MLR), SVR, Quantile Regression (QR), **LSTM**, LSTM - Bayesian Deep Learning (BDL)	MAE (kWh) MLR - 4.829, SVR - 4.077, QR - 4.649, LSTM - **3.627**, LSTM-BDL - 2.782	[63], 2022
30,000 charging session - ACN dataset	Pasaedena, California	ML: KNN, DT, RF, SVR; DL: **LSTM**, GRU, CNN-LSTM, CNN-GRU	MAE: 9.04, 9.02, 9.04, 9.04; MAE: **14.7**, 14.61, 5.43, 5.93	[64], 2023
30,000 charging session - ACN dataset	Pasaedena, California	**LSTM**, RNN - For Timestep: 1, 5, 10	Timestep: 1 - MAE: **5.896**, 8.763; Timestep: 5 -198.29, 73.0294; Timestep: 10 - 212.5277, 133.7053	[59], 2023
New Energy EV charging stations data with 72 charging files with PJM consumption data	China	LSTM-Encoder	MAPE: 1.637, RMSE: 2.267	[65], 2023
This work	**California**	**ML: SARIMA, DL: LSTM, Classification: DT**	**SARIMA - MAE: 7.73, LSTM - MAPE: 2.89, 3.13; Accuracy - DT: 88.8, 89.7**	

1.3 Research Objective

The aim of this work was to investigate the applicability of ML methods for analyzing a historical charging session dataset recorded from 2018-2021, from a real-time charging station located at a university campus garage in Pasadena, California. The objectives of this study are as follows:

1. **Short-term Load Forecasting**: Precise 20- and 30-minutes ahead short-term load forecasts computed based on the charger's connection time and the amount of energy delivered. These predictions are computed for both the aggregated weekdays and weekends using the historical charging session dataset recorded from 2018-2021.

2. **Idle Time (IT) estimation**: Based on the duration spent by the users at the charging station, this IT variable is estimated for the aggregated weekdays and weekends, aiming to classify the likelihood of number of sessions/users available for V2G scenario.

3. **Prediction of Excess kWh**: The excess kWh is computed for the same use case, aggregated weekdays and weekends data based on the difference between the energy delivered and the energy requested by the user from the charging station. Furthermore, the number of users eligible to shed energy back to the grid based on this excess kWh is calculated.

4. **User Behavior Prediction**: The level of uncertainties in the data is determined by analyzing the occurrence of charging sessions during the years 2020-2021. Based on the connection and disconnection time of the chargers, a clustering method is applied to group the number of users and estimate the frequency of charge occurrence. Further, the charging session duration is predicted on a day-ahead scale using ML models.

Paper Organization

The introduction of the research area is stated in Chapter 1, also providing the statistical data of growth of charging infrastructure in the U.S. and research objective. Chapter 2 provides the methodology and results obtained in short-term load forecasting and idle time estimation from the charging station. In Chapter 3, the prediction of excess energy and aggregated user input for V2G services are discussed. Chapter 4 studies the impact of COVID-19 pandemic on the charging station and predicts the EV User behavior based on the charging occurrence. Chapter 6 provides conclusion and future work.

Chapter 2

Short-Term Load Forecasting and Idle Time Estimation using ML Methods

2.1 Introduction

Increase in number of EV sales encourages the user to make use of the public charging stations more frequently. With the increase in number of EV users, the energy management in the electric utilities could be complex which causes load demand. Hence, there is a need for accurate minute ahead EV charging demand forecasting. This chapter focuses on solving two fundamental problems, namely: 1) accurate minute-ahead (20 minute & 30 minute intervals) load forecasts for EVs using four years of historical data sets from a university campus charging station; and 2) probability estimation of aggregated available capacity for V2G connection which could be sold back to the grid as a service. For the problem 1, load forecasts (kWh) from the charging stations are investigated using a series of models such as Seasonal Auto-Regressive Integrated Moving Average (SARIMA), Random Forest (RF), and Neural Networks (NN). For the problem 2, an Idle Time, a parameter which is tracked from the time spent by the EV users (after fully charged) on the charging station. Based on this parameter, the number of users for V2G services are estimated using logistic regression and linear support vector classifier.[1]

[1]This chapter is a slightly modified version of our published paper, P. Rajagopalan, J. Thornby, and P. Ranganathan, "Short-term electric vehicle demand forecasts and Vehicle-to-Grid (V2G) idle-time estimation using machine learning," in 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC), pp. 1279–1286, IEEE, 2023.

2.2 Data Preparation

A real time EV charging station dataset recorded from a university campus garage located in California, obtained from the Adaptive Control Network (ACN) portal was used in this work to investigate the use of EVs on the Caltech campus. This dataset contains 30,000 charging sessions recorded over four-years period from 2018-2021. The important parameters from the dataset considered for this study includes the timestamps such as connection time, disconnection time, charge completion time, and kWh delivered as load parameter recorded from the campus garage [66].

2.2.1 Data Segmentation

The variables from the dataset, such as connect time (indicating the connection time of the charger) and kWh delivered (indicating the load consumed by users per session at the charging station), are considered. These variables are utilized for forecasting the load delivered by users at the charging station. The load delivered data is aggregated to historical weekdays and weekends data separately and scaled to a 24-hours power series. For further analysis, the probability estimation of aggregated capacity of EV users to connect back to the grid were studied, hence the periodical split of kWh delivered to the charging station users are analyzed as shown in Figure 2.1. Since the data collected on Saturday and Sunday are insufficient, they were consolidated and considered as weekend altogether, whereas weekdays were considered individually throughout this study.

The difference between the disconnect time (indicating the disconnection time of the charger) and the done charging time (indicating the actual completion time of charge by the user) were formulated as idle time (IT). This parameter was used for the estimation of number of users connecting back to the grid.

Initially, the overall idle time spent in the charging station on each day collectively for the historical data are studied as shown in Figure 2.2. to analyze which day had more capacity to contribute for V2G services. This analysis clearly indicates that historical Wednesdays followed by Thursdays and Tuesdays, have more number of users to connect back to the grid.

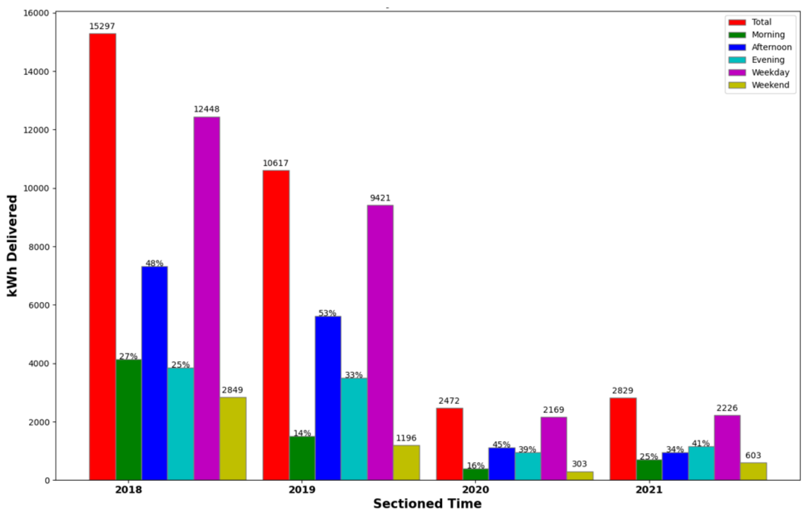


Figure 2.1: kWh Delivered at the Charging Stations from 2018-2021.

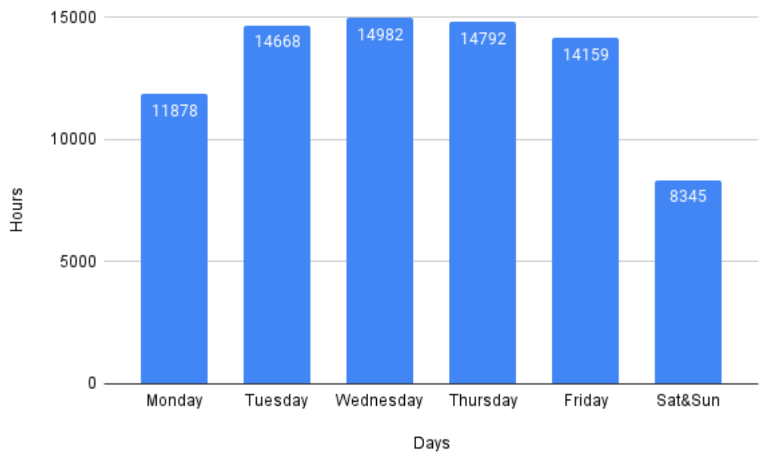


Figure 2.2: Overall Idle Time Recorded at the Charging stations from 2018-2021.

2.3 Methodology

This work presents different ML models to perform the short-term load forecasting. It utilizes the connect time of EV users and kWh delivered from each charging sessions recorded over the historical period of 2018-2021 to create a 24-hour power series at the minute scale for hourly forecasting. This historical data is used to build the model, which in turn can be used to predict the load demand.

2.3.1 SARIMA

SARIMA was originally developed by Box and Jenkins and it is one of the popular time series forecasting method that considers seasonality patterns [67]. An ARIMA model contains of three parts: autoregression p, the degree of difference, and the order of moving average q. The SARIMA model is specified as SARIMA (p, d, q) × (P, D, Q)s, where p, d, and q refers to the orders of the Auto Regressive (AR), differencing, and moving average (MA) parts of the model [68]. The Augmented Dickey Fuller (ADF) test is performed in Python using the statsmodel package through the adfuller function in statsmodels.tsa.stattools [69], the Dickey-Fuller test equation is expressed as:

$$y_t = C + \beta_t + \alpha_{t-1} + \phi_1 \Delta Y_{t-1} + \phi_2 \Delta Y_{t-2} + .. + \phi_p \Delta Y_{t-p} + e_t \tag{2.1}$$

where y_t is the value of the time series at time t. y_{t-1} is the lag 1 of the time series and ΔY_{t-1} is the first order difference of the series at time $(t-1)$. To check for the stationarity of the time series, ADF statistic is used for all weekday and weekend periods [70]. If ADF is more negative, the hypothesis gets rejected, and there exist the presence of a unit root. In this data set, there was a unit root (p > 0.05), and hence data exhibited non-stationarity. To make the data stationary, first order differencing operation was performed. This was performed to make the time series stationary.

2.3.2 Random Forest (RF)

RF is a decision tree-based ensemble learning technique that can be used for creating robust and accurate prediction models [71]. The RF model is based on the bagging concept, where each tree is trained on a random subset and outputs as many decision trees as possible with independent predictions. This method works by sampling the data set at random and then creating a decision tree on each sample. The final predictions are obtained by averaging the predictions of individual trees which helps in reducing over-fitting and noise. Consider a train set $S = (X^m, Y^m)^m_{(M=1)}$, where $X \in R^D$ which contains the input features with parameters such as connection time of the charger and energy delivered based on the request by the user. M is the amount of samples with the subset S_t for the training set S. Multiple iterations are conducted on the input data to generate bootstraps for each tree. The prediction of all the independent trees are averaged as a single aggregated value, once after the training and testing phases are completed on the bootstrap data.

2.4 Short Term Load Forecasting: Results and Discussion

This section highlights the hyperparameters of the ML models, accuracy metrics and their findings for forecasting the kWh delivered from the charging station. Figure 2.3. illustrates the minute level kWh delivered from historical Mondays data recorded from 2018-2021. The prediction of kWh delivered in the charging stations are estimated using the error metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) as mentioned in (2.2) and (2.3). The visualization of models predicting over the test data and it's corresponding forecasts are shown in Appendix Figure C1, C2 and C3 respectively.

$$MAE = \frac{1}{n} \sum_{i=1}^{D} |y_i - \hat{y}_i| \tag{2.2}$$

where y_i is the observed value of the ith observation, n is the number of samples. $\hat{y}_i$ is the predicted value of the observation.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{2.3}$$

Table 2.1 explains the hyperparameter values used for the ML models. It utilizes the connect time of EV users and kWh delivered from the charging stations over the historical time frame of 2018-2021 to create a 24-hour power series at a minute-by-minute scale for minute level forecasting. The historical data is used to build the models, which in turn can be utilized to predict the load delivered.

Table 2.1: Hyperparameters for RF, SARIMA and NN Models.

ML Models	***Parameter Description***	***Values***
RF	n_estimators	100
	random_state	5
SARIMA	p, d, q	(1,1,1)
NN	activation	ReLu
	dense	64
	optimizer	Adam

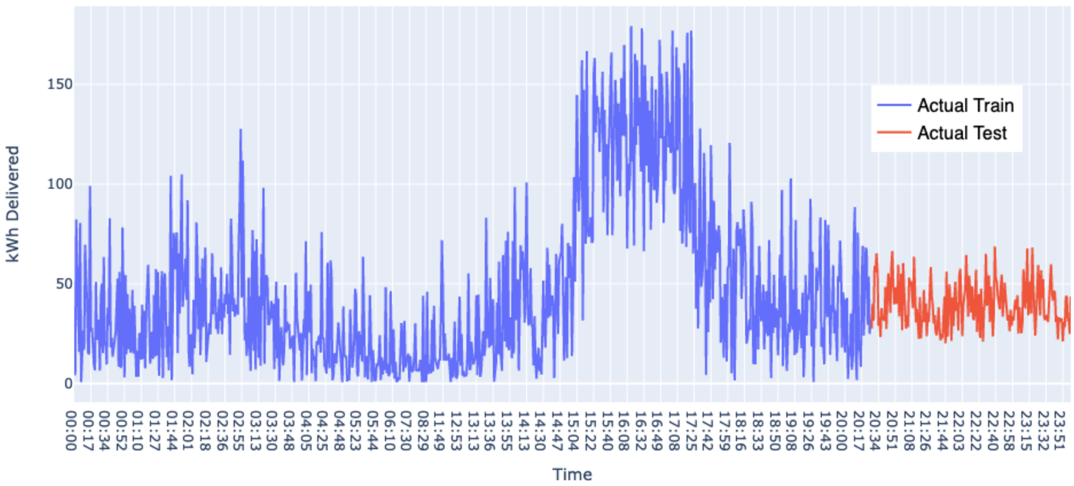


Figure 2.3: kWh Delivered on Historical Mondays on a 24-hours Scale.

Observation from Table 2.2 shows that the RF model performed better, with a MAE score of 7.26 for Mondays, 9.79 for Wednesdays and 9.16 for Fridays data respectively. SARIMA model achieved the second-best result, with MAE of 7.73 for Mondays, 10.48 for Tuesdays and 9.95 for Fridays respectively. This is due to the ability of the ensemble method, RF model handles the minute-level data containing noisy data points with multiple decision trees, making better performing model over SARIMA and NN. For the purpose of explanation, the prediction results of historical Mondays are illustrated in Figure 2.4 showing the prediction of ML models, SARIMA, NN and RF over the test data.

Table 2.3 shows RMSE scores for the differed ML models. The performance of the models is similar in both MAE and RMSE, where RF model performed

Table 2.2: Error Metrics - MAE for Short term Load Prediction Using Different ML Models.

Models	Days					
	Mon	***Tues***	***Wed***	***Thurs***	***Fri***	***Sat & Sun***
SARIMA	**7.732**	11.858	**10.483**	10.165	**9.954**	12.560
NN	8.425	13.113	11.076	13.11	10.532	12.787
RF	**7.263**	11.754	**9.796**	9.980	**9.163**	11.276

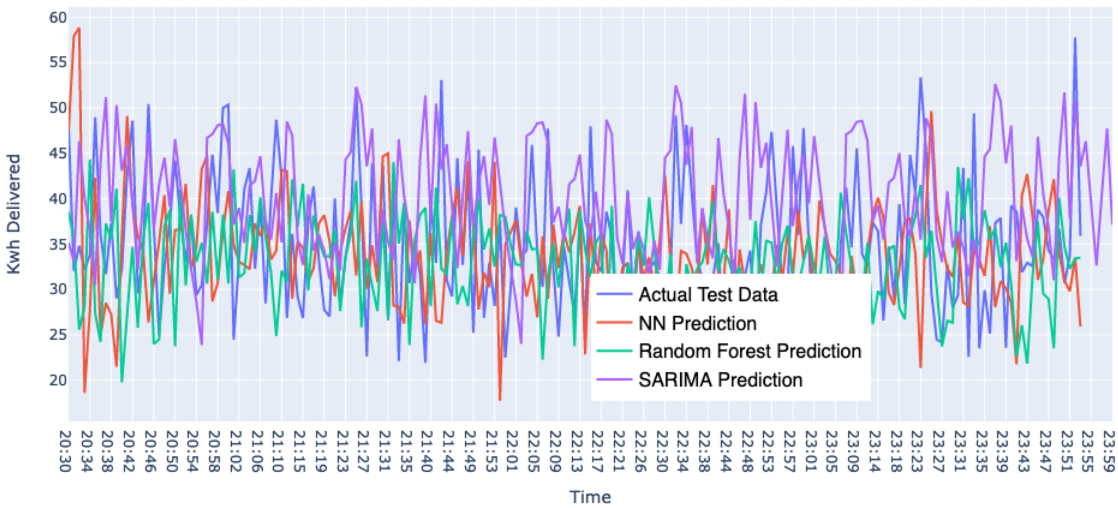


Figure 2.4: Prediction of kWh Delivered using Different ML Models for Historical Mondays.

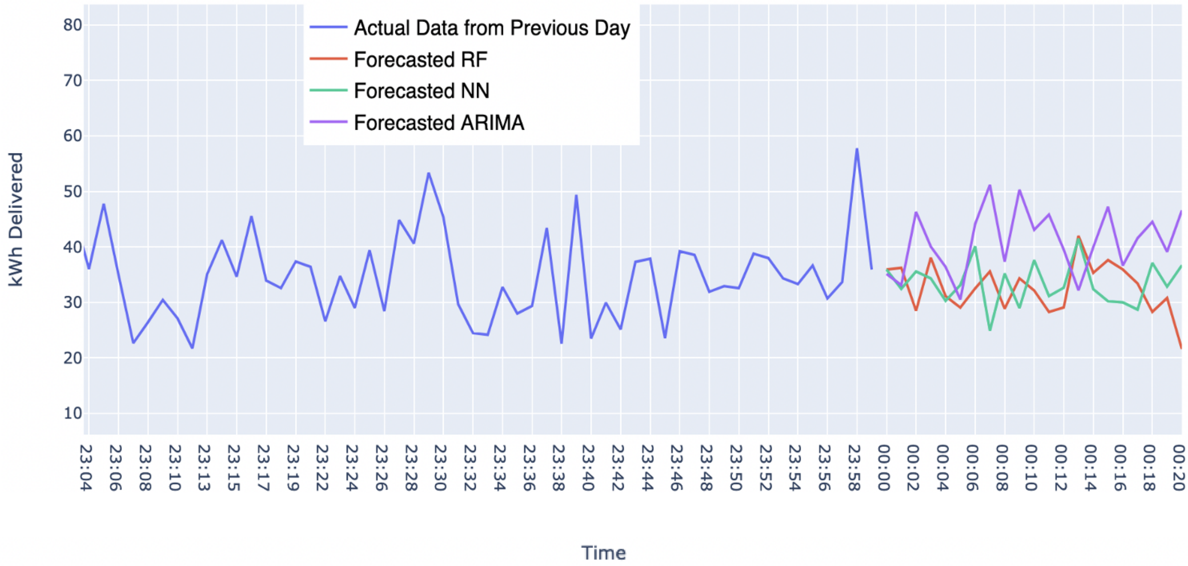
better considering the data from historical Mondays (9.15), Wednesdays (12.64) and Fridays (12.56). In the previous case, the historical data from 2018-2021 is considered for the short-term load forecasting study. Further, this study is extended to year wise segmentation. Therefore, the data from 2020-2021 having less charging sessions occurred, due to COVID-19. This data is consolidated together to examine the performance of each ML models. The performance metrics of each ML models for this cases are as shown in Table 2.4. In this extended result, RF performed better with respect to MAE, which agrees with the previous case. Figure 2.5 represents the 20- & 30-minutes ahead load forecasting results for historical Mondays from 2018-2021.

Table 2.3: Error Metrics - RMSE for Short term Load Prediction Using Different ML Models.

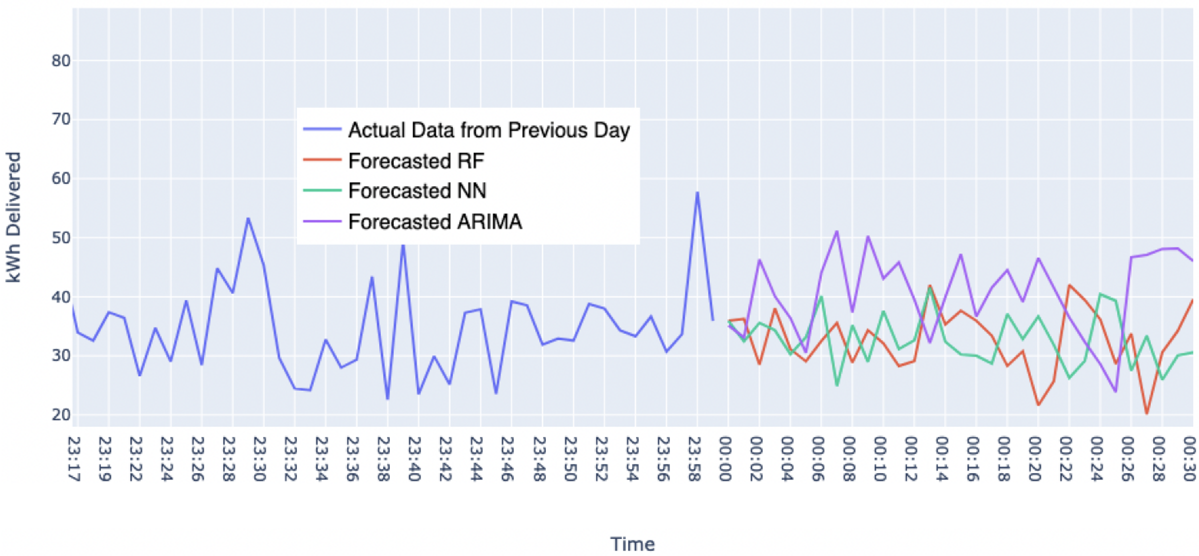
Models	Days					
	Mon	***Tues***	***Wed***	***Thurs***	***Fri***	***Sat & Sun***
SARIMA	10.812	15.125	12.133	11.501	13.479	13.729
NN	9.736	16.293	13.611	16.293	13.575	15.467
RF	**9.156**	15.261	**12.640**	13.085	12.569	15.474

Table 2.4: Error Metrics - MAE for Short term Load Prediction Using Different ML Models.

Models	Year	Days					
		Mon	***Tue***	***Wed***	***Thu***	***Fri***	***Sat&Sun***
RF	2018	5.698	5.224	5.224	6.783	5.158	6.533
	2019	9.104	6.626	6.753	6.605	7.026	7.940
	2020-2021	7.830	6.262	8.203	8.238	9.368	9.788
NN	2018	4.934	5.097	5.097	8.005	5.509	6.471
	2019	10.419	7.083	6.709	6.511	6.520	9.674
	2020-2021	13.529	10.360	9.068	8.799	8.810	13.290
SARIMA	2018	6.485	5.398	5.314	5.909	5.246	6.951
	2019	7.572	5.911	6.592	6.743	7.326	7.826
	2020-2021	7.032	5.725	5.823	7.276	8.625	9.522



(a) 20 Minutes Ahead Forecasting.



(b) 30 Minutes Ahead Forecasting.

Figure 2.5: Load Forecasting for Historical Mondays from 2018-2021.

2.5 Estimation of Number of Users based on Idle Time

Logistic Regression (LR): As the historical data (Figure 2.1) shows that kWh delivered varied from 10 kWh to 65 kWh, this data set is good to explore, for estimating the number of users through Idle time evaluation (after fully charged, not charging) for potential V2G services. For our simplicity, we chose 30 kWh as threshold value to look for users who have at least charged to this level, and analyze them by weekdays/weekends. A sigmoid function was used within logistic regression method, as this function can map any real-valued inputs between 0 and 1 to estimate the probabilistic instance classification. The probability estimation using logistic regression is calculated as shown in equation (2.4) and (2.5). This logistic model predicts the output of the variable as binary (0 or 1), based on the threshold of users who either consumes <30kWh or >30kWh respectively. To determine the aggregated number of users, the daily historical data is split into following hourly groups for ease of visualization: 12am to 5am (group 1), 6am to 11am (group 2), 12pm to 5pm (group 3), 6pm to 11pm (group 4) and so on. The classification accuracy is shown in Table 2.5 which indicates group 3 has good number of users.

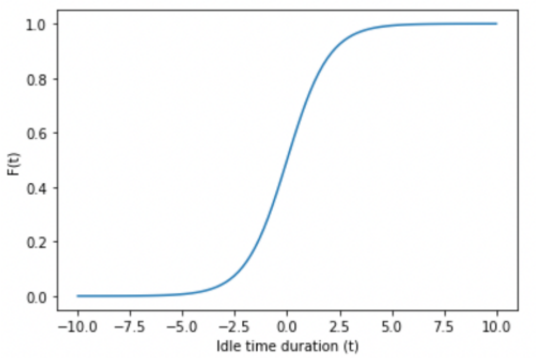


Figure 2.6: Sigmoid Function Curve to Determine Idle Time.

$$\hat{p} \ = \ h_\theta(x) = \ \sigma(X^T\theta) \tag{2.4}$$

where the logistic function form is,

$$\sigma(t) = \frac{1}{1 + exp(-t)} \tag{2.5}$$

$if \ \ t < 0, \sigma(t) < 0.5;$

$if \ \ t >= 0, \sigma(t) >= 0.5;$

Hence the regression is obtained using $\hat{p}$ that x belongs to class 1 ($0 \leq \hat{p} \leq 1$).

The predict $\hat{y}$ is set with,

$$\hat{y} = \begin{cases} 0 & if \;\; \hat{p} < 0.5 \\ 1 & if \;\; \hat{p} \geq 0.5 \end{cases}$$

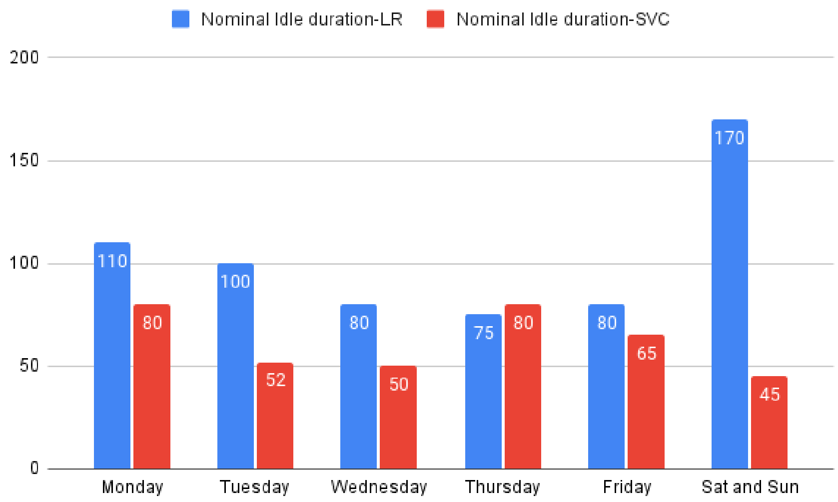


Figure 2.7: Nominal Idle Time Duration for V2G Services Using Logistic Regression and Linear SVC for group 3.

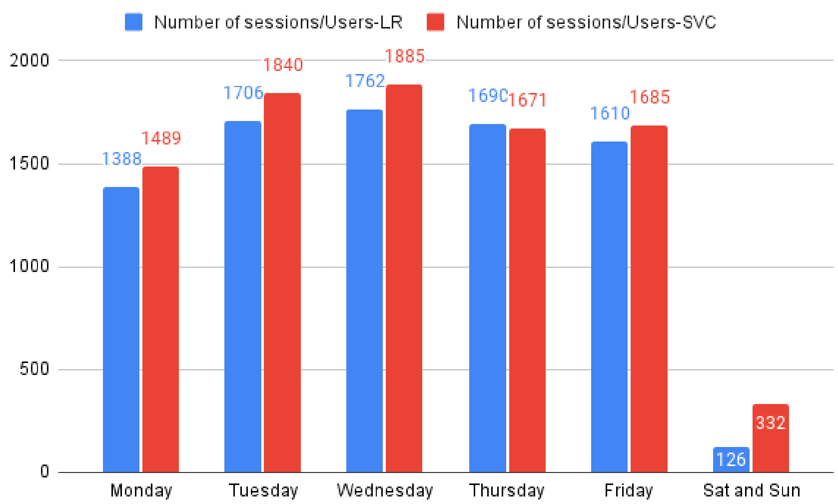


Figure 2.8: Prediction of Aggregated Number of users Available for V2G Services using Logistic Regression and Linear SVC for group 3.

These values are compared to the nominal idle time evaluated from the sigmoid function as shown in Figure 2.6 in logistic regression. The result shows that the hour 12 to 17 which indicates 12pm to 5pm are having a greater number of users on daily basis as shown in Figure 2.7 with respect to the idle time predicted using classification models.

Table 2.5: Accuracy Metrics for LR and Linear SVC Models.

Accuracy	**Days**					
	Mon	***Tue***	***Wed***	***Thu***	***Fri***	***Sat&Sun***
LR	81.83	84.36	87.97	84.53	86.46	72.10
Linear SVC	85.08	91.06	91.81	87.05	86.82	94.84

Linear Support Vector Classifier (Linear SVC): This model creates a hyperplane or series of hyperplanes in a high or infinite dimensional space for classification. The common attributes used between the LR and the Linear SVC are *predictproba*, $Coef$, and Intercept under the large linear library classification called LIBLINEAR [72]. Similar to LR method, the threshold of aggregated available capacity of users is set to the sessions that occurred beyond 30kWh. A binary classification using SVC method is used to estimate the probability of the instances. Using the mathematical formulation from the equation (2.6), likelihood of a charging station user connecting back to the grid are estimated with the criterion of probability that falls over 50%. The result shows that the hour 12 to 17 which indicates 12pm to 5pm are having a greater number of users on daily basis as shown in Figure 2.7 and Figure 2.8.

$$\min_{w,b} \frac{1}{2}(w^T w) + C \sum_{i=1}^{n} max(0, 1 - y_i(w^T \Phi(x_i + b)) \tag{2.6}$$

where ϕ is the identity function supported by LinearSVC, w is the linear classifier weight vector, x_i and y_i are the training variables from the i th set, C is the kernel weights, b is the constant specified for the model.

The discrete values are categorized into 0 or 1 based on the threshold of users belonged to the category less than 30kWh and greater than 30kWh respectively.

These numbers are compared to the nominal idle time calculated from the parameters are set with *random state* = 10 and *max iteration* = 1000. The accuracy values for SVC are shown in Table 2.5. SVC, on the other hand, are also in agreement with LR that group 3 has significant users who exceed 30kWh consumption and may be available for V2G services.

Although the number of users predicted by the logistic and linear SVC model are not exactly the same, we took average of these models as a measure to proceed further. Hence, the average number of users in group 3 time slot for Mondays-Weekends are as follows: 1438 (Mon), 1773 (Tues), 1823 (Wed), 1680 (Thu), 1647 (Fri) and 229 (Sat/Sun: Weekends). These users are likely to connect back to the grid service.

2.6 Conclusion

This chapter has utilized SARIMA, RF and NN to perform short-term load forecasts with respect to connection time and kWh delivered from the EV charging station. The preliminary result indicates that all of the three models do agree that Monday's forecast yield a lower MAE values, while Tuesdays, Thursdays, and weekends contain large errors. The chapter also investigated the estimation of number of users available for V2G service based on Idle time parameter. This was achieved using SVC and Logistic regression models. The results indicate that historical Wednesdays were having more number of users (1823) for connecting back to the grid. Such V2G capacity estimates can be useful to aid in any charging infrastructure supply/demand imbalances and to assist in querying these users on their willingness for V2G service.

Chapter 3

Prediction of Excess Energy from the Charging Stations

3.1 Introduction

As Carbon dioxide emissions have steadily raised over the last two decades, it became a significant concern in the context of climate change. According to the U.S. Environmental Protection Agency (EPA), total Carbon dioxide emission in 2021 is 6,340 million metric tons, among which transportation sector contributes up to 28% [73]. EV presents a great opportunity in reducing this level of emission and also contributing in demand response through V2G services. Hence, the study of EV charging forecast is an important factor in managing the charging stations [74]. The prediction of excess energy available from a charging station provides potential benefits in efficient management of the charging facilities. This chapter addresses two primary objectives: (1) predicting excess kWh available from a real-time charging station; (2) determining the number of users corresponding to 15% and 30% of excess kWh. For the scenario 1, multiple ML models such as SVR, GBR, LSTM, and RF are deployed. There are several missing values in the dataset, these missing values of the excess kWh are imputed using Pandas data frame interpolation for robust training. For the scenario 2, the estimation of users availability for V2G services using DT and KNN models are performed.[2]

[2]This chapter is a slightly modified version of our published paper, P. Rajagopalan and P. Ranganathan, "Predicting Excess Energy and Estimating Users for Vehicle-to-Grid (V2G) Services Using Machine Learning," (accepted in Ubiquitous Computing, Electronics and Communication Conference, 2023).

3.2 Data Pre-processing

This chapter utilizes the same dataset from the previous chapter. However, the parameter consideration varies for this work. The important parameters from the dataset considered for this study includes the timestamps such as connection time, and load parameters such as kWh requested, kWh delivered recorded from the campus garage [66].

3.2.1 Data Interpolation with Pandas

The difference between the energy delivered and the energy requested by the users from the data frame is used to calculate the excess kWh from the charging station. The values less than zero from the excess kWh formulation are considered missing values, and these missing values are imputed using Pandas data frame interpolation as time based data points with respect to the available excess kWh and the connect time of each session. These values are considered for setting up the training sample set. Interpolation involves in estimating and adapting a method to generate new data values within a given series of definite, pre-existing data points [75]. This dataset was used to forecast the amount of excess kWh generated at a charging station for each day of the week. The collected data is divided into two separate datasets: one for historical weekdays and other for the historical weekends. This enables more precise and targeted predictions of the amount of excess kWh generated at the charging station on any given day. Figure 3.1 represents the kWh requested on each day for 2018 to 2021.

The parameters considered from the dataset such as connect time signifies the duration of the charger connected to a single EV, while excess kWh available is the aggregated amount of excess energy that is available at the charging station.

To scale the excess kWh data to a 24-hour power series, it is aggregated into distinct historical weekday and weekend data. This is accomplished by mapping the excess kWh data to the corresponding days in the historical data and subsequently using an interpolation algorithm to bridge the gaps. This approach ensures the precision of excess kWh data in the 24-hour power series.

This same set of data frame can be used to determine the aggregate number of users potentially participating in V2G services. By analyzing the data frame, the number of users qualifying for V2G services when considering 15% of excess kWh are estimated. Similarly, the qualified number of users when it increases to 30% of excess kWh. This data can be used to gain a better understanding of the potential user pool for V2G services. Figure 3.2 illustrates the Step-by-Step Approach for creating the forecasting model using different ML techniques.

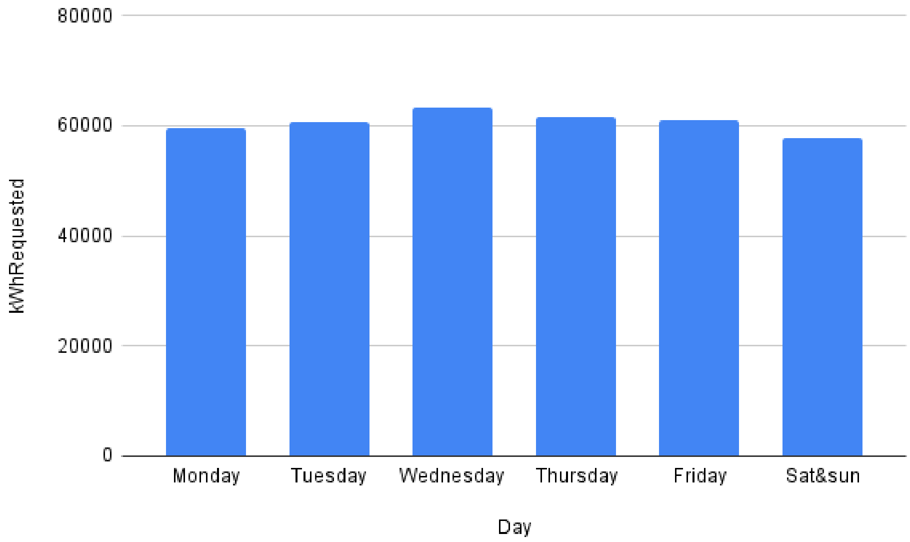


Figure 3.1: kWh Requested by the Users on Each Day from 2018-2021.

3.3 Methodology

This work presents different ML models for predicting the aggregated excess energy from the charging stations. It utilizes the connect time of EV users and excess kWh from each charging sessions recorded over the historical period of 2018-2021 to create a 24-hour power series. This series contains the minute level scale of data for hourly forecasting. This historical data is used to build the model, which in turn can be used to predict the excess energy.

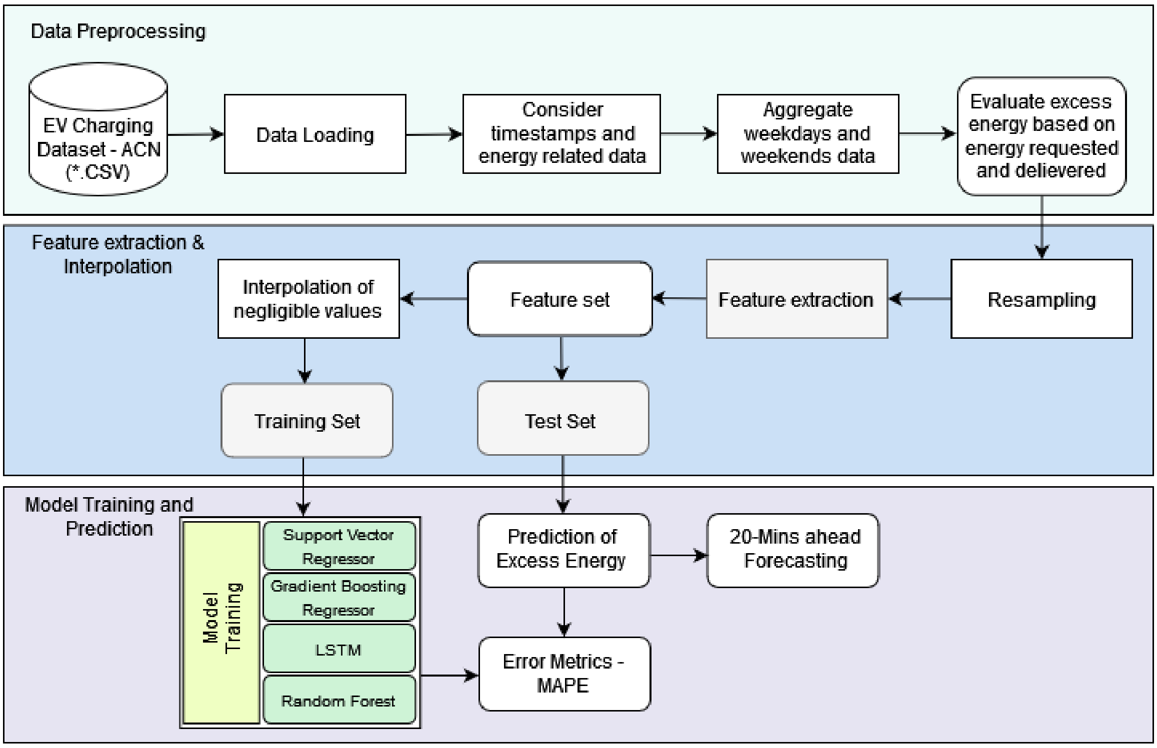


Figure 3.2: Step-by-Step Approach for creating the Forecasting Model.

3.3.1 Support Vector Regression (SVR)

The SVR is a ML algorithm proposed by Cortes, that relies on the principle of reducing structural uncertainty in order to better generalize a fewer amount of samples [76]. This method generates an optimal hyperplane with a decision boundary, one of which achieves the largest margin among two classes. These hyperplanes are structured with support vectors, which reduces the over-fitting errors present in conventional prediction models. Hence this model is commonly used in short term load forecasting problems especially in terms of charging, SoC and load consideration [61]. For the purpose of utilizing the SVR model, in this study, a step-by-step process is followed based on the problem formulation from the Scikit sklearn.svm [77].

Consider a training set $\Gamma = \{x_i, y_i\}_{i=1}^{n}$ where $x_i \in \mathbb{R}^d$ and $y_i \in \mathbb{R}$, the goal of SVR model is to generate a prediction model for the foreseen instances. When the dataset Γ which is dependent linearly, the SVR solves the problem with the below equations,

$$min_{\omega,b,\xi,\xi_i^*} \ \frac{1}{2}\omega^T\omega + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{3.1}$$

$$s.t. \begin{cases} y_i - (\omega^T x_i + b) \leq \varepsilon + \xi_i \\ (\omega^T x_i + b) - y_i \leq \varepsilon + \xi_i^* \\ \xi_i \xi_i^* \geq 0 \end{cases} \tag{3.2}$$

where ω denotes the weight vector, and b represent the bias, and x_i and y_i, $i = 1..N$ are training samples. ε is the maximum value of the error, ξ and ξ^* denotes the distance between actual values, and related boundary values of ε-tube, $C \geq 0$ marks the trade-off of training error. The constraint condition of the above mathematical model is to accurately predicts all training samples from the interpolated excess kWh data. The kernel function of the SVR are linear, tanh, polynomial and Gaussian or Radial Based Function (RBF) [78].

The linear kernel is,

$$k(x, z) = x^T z, \tag{3.3}$$

the tanh kernel is,

$$k(x, z) = tanh(gx^T z + c), \tag{3.4}$$

the polynomial kernel is,

$$k(x, z) = (x^T z + c)^d, \tag{3.5}$$

and the gaussian or RBF kernel is,

$$k(x, z) = exp(\frac{-(x - z)^2}{2 \times \delta^2}), \tag{3.6}$$

where g denotes the slope of the kernel, c represents the offset of the polynomial, d is the degree of the polynomial kernel, δ is the width of gaussian kernel. In this study, the RBF kernel is used to implicit the feature mapping.

3.3.2 Gradient Boosting Regression (GBR)

GBR is an ensemble ML technique focuses on boosting method that generates base models sequentially. The general idea of this model is to sequentially train the weak learners within the decision tree and adjust their weights to improve the overall prediction [79]. The developed models in sequence improves the prediction accuracy by emphasizing training set that are tedious to estimate. The modified GBR method using a regression tree of fixed size base model is proposed by Friedman et al. to improve the quality of the model [80].

In this study, the modified GBR was used for the short-term load forecasting with respect to the connection time of the charger and energy delivered based on the request made by the user. Considering the number of leaves for each tree is J. Each tree segmented the input into J splitted regions as $R_{1m}, R_{2m}..R_{jm}$ to predict the values of b_{jm} for R_{jm} region. The regression model is defined as:

$$g_m(x) = \sum_{j=1}^{J} b_{jm} I(x \in R_{jm}) \tag{3.7}$$

where

$$I(x \in R_{jm}) = \begin{cases} 1, & if x \in R_{jm} \\ 0, & otherwise \end{cases} \tag{3.8}$$

Utilizing the regression tree to replace $g_m(x_i)$ in the gradient boosting method, the model equation is represented as,

$$f_m(x) = f_{m-1}(x) + \rho_m g_m(x) \tag{3.9}$$

$$\rho_m(x) = argmin_\rho \sum_{i=1}^{n} L(y_i, f_{m-1}(x_i) + \rho g_m(x_i)) \tag{3.10}$$

which is modified into,

$$f_m(x) = f_{m-1}(x) + \rho_m g_m(x) \tag{3.11}$$

The model mandates to determine the optimal number of iterations or M to minimize risks leading to fitting the model closely to the training data and to the poor generalization ability. To avoid the over-fitting, the number of gradient boosting iterations are controlled, by scaling the factor $J \in (0, 1)$. Then the equation becomes,

$$f_m(x) = f_{m-1}(x) + J. \sum_{j=1}^{J} \rho_{jm} I(x \in R_{jm}) \tag{3.12}$$

where J denotes the learning rate, which controls each model with a factor $0 \leq J \leq 1$. The tradeoff between the iterations requires a small value of J, with larger value of M to obtain better training model.

3.3.3 Long-Short Term Memory

LSTM is a distinctive version of the RNN. This model has a memory structure capable of storing and recalling data correlation details in a time series, which can adapt to both long- and short-term load forecasting [81]. In this case, the model is utilized for the short-term excess energy prediction. This is accomplished by bringing memory state units C_t, as well as three gate units to the implicit layer. The three gate units, input, output, and forget gates regulate the flow of minute wise data in terms of excess kWh. Considering the LSTM layers, the input gate regulates the inflow of data into memory units, while the output gate regulates the outflow of information out of memory units. The forget gate measures the information maintained in memory units to control the use of historical sequential data as a 24-hour scale information [82].

To make predictions using this model, The input series x_t, the hidden state h_{t-1} at time state t_1, and the cell information C_{t-1} which includes the forget gate f_t, input gate i_t, and output gate o_t collectively determine the output h_t of the LSTM-RNN cell at time t.

The forget gate f_t regulates the memory details that were erased from memory C_{t-1} in the preceding time step. The updated details from the new input data is controlled by the input gate, and the memory C_t detail is scaled by the output gate. These components share a common structure. The equations governing these processes are as follows,

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \tag{3.13}$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_j) \tag{3.14}$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \tag{3.15}$$

The new cell information or memory C_t is derived using the derived outputs of the forget and input gates,

$$C_t = tanh(W_c x_t + U_o h_{t-1} + b_c) \tag{3.16}$$

$$C_t = f_t C_{t-1} + t_t C'_t \tag{3.17}$$

W_c and U_c represent the weight matrices associated with memory, b_c is the corresponding bias. where σ represents the sigmoid function. Each of the three gate units relies on the present input x_t and the preceding output h_{t-1}. The notations W, U, b refer to the weight matrix for input, memory, and bias, respectively. Parameters for the forget gate, input and output gates are denoted by the subscripts f, i, o respectively.

3.4 Excess Energy Prediction: Results and Discussion

This section highlights the hyperparameters of the ML models, accuracy metrics and findings for predicting the aggregated excess energy from the charging station. Table 3.1 explains the hyperparameter values used for the ML models. It utilizes the connect time of EV users and excess kWh in the charging stations over the historical time frame of 2018-2021 to create a 24-hour power series at a minute-by-minute scale for hourly forecasting.

The prediction of excess energy utilized by the users corresponding to the connecting time captured from the charging stations can be accurately computed using the MAPE as mentioned in equation 3.18.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}|\frac{y_i - y_p}{y_i}| \tag{3.18}$$

where n denotes the samples present in the test set, y_p represents predicted values and y_i signifies the actual short-term load values. In this case, the load values denotes to the excess kWh from the charging station.

Table 3.1: Hyperparameters for RF, SVR, GBR and LSTM models.

ML Models	***Parameter Description***	***Values***
RF	n_estimators	100
	random_state	5
SVR	kernel	RBF
	gamma	'scale'
GBR	n_estimators	100
	learning_rate	0.1
	max_depth	1
	random_state	5
LSTM	encoder with feature attention mechanism	3 LSTM layers
	batch size	256
	optimizer	Adam
	learning rate	0.001
	epochs	100
	activation Function	ReLu

To compute the accuracy of the prediction models, MAPE metric evaluation is used in this study, Table 3.2 indicates the number of data samples on each days and associated number of days aggregated for the period of 2018-2021.

Observation from Table 3.3 shows that the LSTM model performed better, with a MAPE score of 3.13 for Mondays, 4.21 for Tuesdays and 3.37 for weekends data respectively. RF model achieved the second-best result, with MAPE of 4.67 for Mondays, 4.81 for Tuesdays and 4.98 for weekends respectively.

Table 3.2: Number of Samples Considered for ML Models from 2018-2021.

Data Parameters	**Days**					
	Mon	***Tues***	***Wed***	***Thurs***	***Fri***	***Sat & Sun***
Number of samples	2650	2923	2973	2769	2854	2219
Days Aggregated	155	155	154	152	156	311

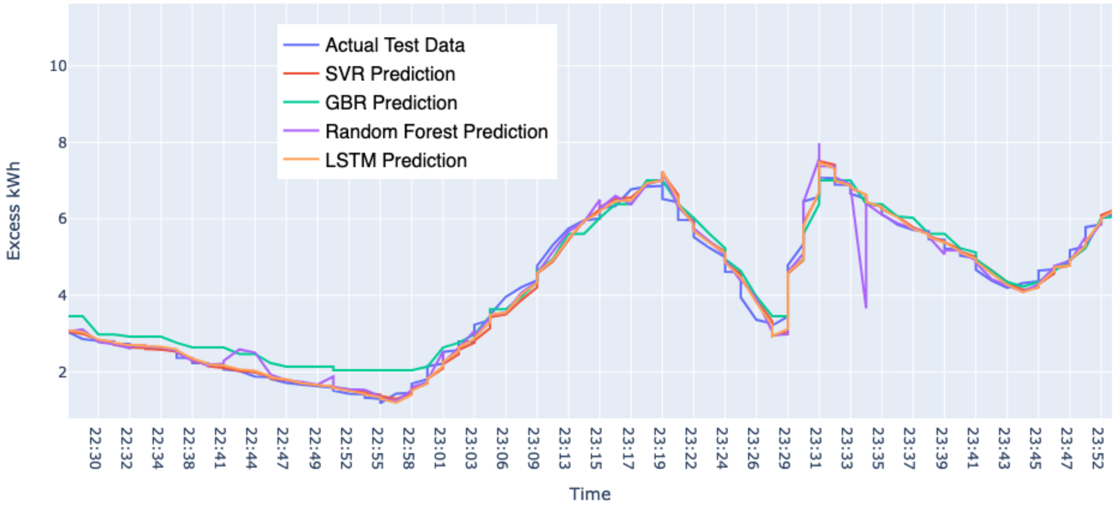


Figure 3.3: Prediction of Excess kWh for Historical Mondays within the Data Period of 2018-2021.

The SVR model achieved MAPE score of 3.87 for Wednesdays and 3.906 for Thursdays with its highest of 12.031 for historical Friday data, this scale is repeated in GBR and RF. It is also notable that the GBR model performed poorly with MAPE score of 11.45 for Mondays, 5.84 for Tuesdays and, 9.31 for weekends data respectively. Additionally, MAPE obtained by the models for historical Wednesdays consistently remains low, primarily due to the larger number of samples in the training dataset.

Figure 3.3 illustrates the prediction of ML models, SVR, GBR, RF and LSTM over the test data. For simplicity, the historical Mondays prediction is shown here. The visualization of prediction of ML models, SVR, GBR, RF and LSTM over the test data for Mondays-Weekends are illustrated in Appendix Figure B1.

Table 3.3: Error Metrics - MAPE for Excess energy Prediction Using Different ML models.

Models	Days					
	Mon	***Tues***	***Wed***	***Thurs***	***Fri***	***Sat & Sun***
SVR	4.240	5.410	3.876	3.906	12.031	4.880
GBR	11.451	5.839	4.705	9.959	8.392	9.315
LSTM	**3.137**	**4.215**	3.122	4.753	4.112	**3.378**
RF	4.672	4.819	4.436	3.597	8.247	4.982

Considering the performances of ML models, the least performed GBR and the well performed LSTM are considered again for the data period of 2018-2019, neglecting the year 2020, as the data pattern is irregular and the period of 2021, which is partially available from the dataset. Table 3.4 represents the MAPE score for the excess kWh prediction using the ML models, GBR and LSTM for the period of 2018-2019.

Table 3.4: Error Metrics - MAPE for Excess Energy Prediction Using GBR and LSTM for 2018-2019.

Models	Days					
	Mon	***Tues***	***Wed***	***Thurs***	***Fri***	***Sat & Sun***
GBR	10.264	2.963	4.489	13.322	4.283	8.421
LSTM	**3.512**	3.859	**2.896**	7.926	**3.035**	6.668

This indicates that the LSTM model is the accurate and reliable model for this type of data consideration. When it comes to predicting this type of data with a significant temporal component and complex, non-linear correlations between variables, LSTM models clearly outperform RF and other models. This is because, LSTM models are well-suited to capture the underlying patterns in data, even when the data contains a significant degree of noise or fluctuation. For this type of data, they are more dependable and accurate than RF. LSTM models are also better at modeling extended data sequences, giving a better alternative for larger datasets. As a result, they are frequently deployed in applications like time series prediction. RF, on the other hand, may be better suited for tabular data with more characteristics and simpler and more linear correlations between variables. RF can handle a large variety of data formats and can capture complicated correlations between features.

3.5 Estimation of Number of Users Based on Excess kWh

Decision Tree (DT): DT is a supervised ML algorithm used for classification, where the data is continuously split into different tree nodes that can be used to predict a target value by decision making from the given training data [83]. The algorithm typically selects the most informative feature from the tree and creates a split at that feature value to create two new branches. This will continue to create new branches by repeating the same procedure until all the leaf nodes belong to the same class. The decision tree is then used to make predictions by traversing from the root node to a leaf node. The leaf node contains the class label for the data point. Therefore, decision tree classification can be used for V2G estimation to identify patterns in the data and make predictions about the aggregated number of users from the charging station. The accuracy metrics are used for the evaluation of the DT and KNN classification approach as shown in equation 3.19.

$$Accuracy \;= \sum_{c}^{C} \frac{tp_c}{N} \tag{3.19}$$

where tp denotes the true positives for class c, C is the quantity of classes, and N is number of instances. The performance of the classification models evaluated using the accuracy metrics for 15% and 30% of excess kWh are shown in the Table 3.5.

Preliminary results from the DT model shows that it performed better in determining the number of users. The accuracy of 89.7% for Thursdays and 88.6% for Fridays respectively, when considering 15% of excess kWh are observed. Also, the accuracy of this models slightly dips with 30% of excess kWh consideration. This is due to the quantity of training sample set that goes down with this

Table 3.5: Accuracy Metrics for DT and KNN Models to Estimate the Number of Users Based on 15% and 30% of Excess kwh.

Models	Accuracy in %					
	Mon	***Tues***	***Wed***	***Thurs***	***Fri***	***Sat & Sun***
DT for 15% of Excess kWh	84.5	88.8	84.0	89.7	88.6	85.3
DT for 30% of Excess kWh	68.6	69.5	66.3	68.2	68.6	70.0
KNN for 15% of Excess kWh	79.4	86.8	78.1	85.1	86.0	81.7
KNN for 30% of Excess kWh	62.0	60.1	61.1	60.1	60.3	65.0

threshold.

K- Nearest Neighbors (KNN): KNN is also a supervised machine learning method and is frequently used for classification-based problems which proposes to locate k training instances nearer to the targets in the training set [84].

The neighbors are derived from the object sets from a specific class, hence the selection of k value is important. According to the k value set, the test samples are predicted corresponding to ‘ + ’ KNN rule.

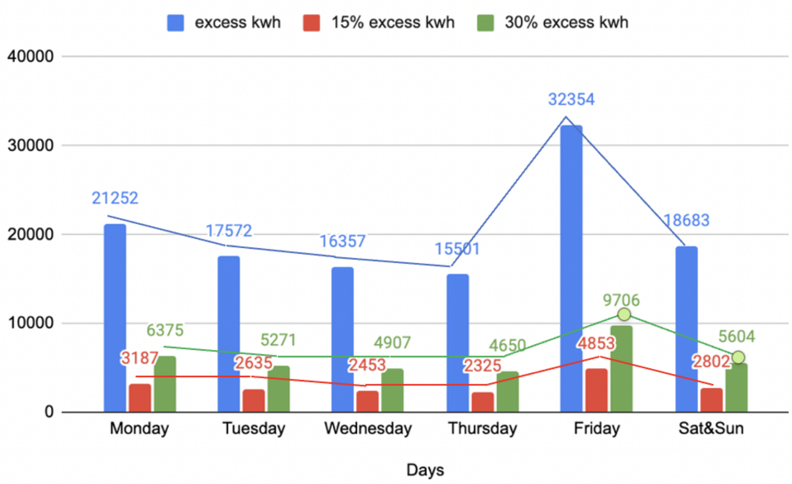


Figure 3.4: Excess kWh from the Charging Station on Weekdays and Weekends for the Data Period of 2018-2021.

This model picks the k nearest samples from the training dataset for a given test sample and makes predictions using a binary classifier [85]. Figure 3.4 represents the aggregated excess kWh and the corresponding 15% and 30% thresholds of excess kWh from the charging station, on weekdays and weekends during 2018-2021.

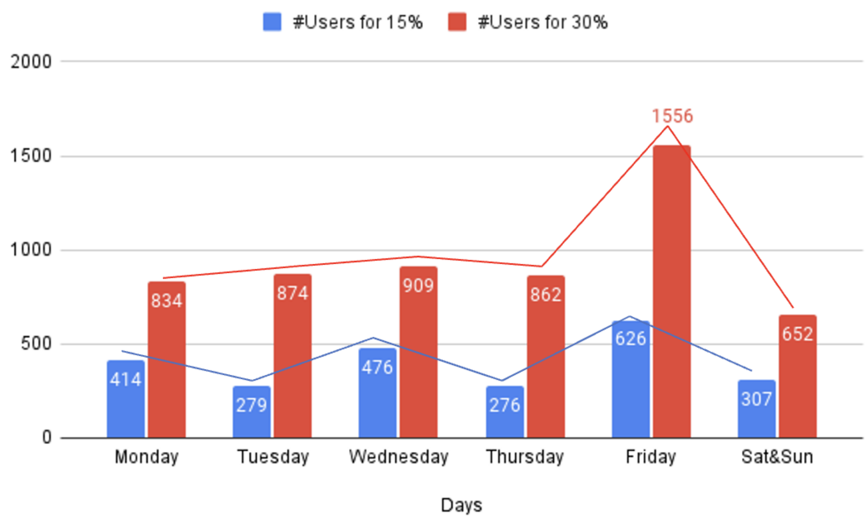


Figure 3.5: Potential Number of Users to Contribute for V2G on Weekdays and Weekends Based on 15% and 30% of Excess kWh.

Figure 3.5 shows the number of users who can contribute for V2G based on the excess kWh computed using DT model. Charging sessions are considered based on the aggregated 15% and 30% excess kWh from historical weekday and weekends data spanning 2018-2021. The results indicate the aggregated excess kWh figures, as high as 32,354 for the historical Fridays through charging events, closely followed by Mondays and weekends. Based on this aggregated excess kWh, number of users are computed for both 15% and 30% thresholds using classification models.

The results indicate the user counts significantly increases on Fridays, reaching 626 users for 15% of excess kWh and 1,556 users for 30% of excess kWh, respectively. However, this pattern of user counts corresponding to the excess kWh is not followed on Mondays and weekends, where the number of users is higher for Wednesdays. This difference is attributed to the increased number of charging sessions recorded on historical Wednesdays as many as 5,491 charging sessions compared to 4,773 on Monday and 4,988 on weekends. Therefore, it is evident that estimating the users cannot be statistically determined solely based on the excess kWh; it can also rely on the number of sessions recorded.

3.6 Conclusion

This chapter compared four ML techniques, SVR, GBR, RF and LSTM, for predicting excess energy using the EV charging dataset. According to preliminary results, LSTM outperformed other models, with MAPE scores of 3.13 for Mondays, 4.21 for Tuesdays and 3.37 for weekends data respectively. GBR, on the other hand, had the highest error rate of all, with MAPE scores of 11.45 for Mondays, 5.84 for Tuesdays and, 9.31 for weekends data respectively. This paper also investigated the potential number of users likely to connect for V2G service based on 15% and 30% of excess kWh available from the charging stations using DT and KNN algorithms. The choice of analyzing 15% and 30% figures of excess kWh available from the charging station is to assess the V2G potential.

Chapter 4

User Behavior Prediction in the Charging Station using ML Models

4.1 Introduction

The global spread of the COVID-19 pandemic has significantly impacted the EV industry. The lockdown restriction has resulted in interruption in the use of public charging infrastructures and travel pattern. In April 2020, the mobility patterns to workplaces were 50% below baseline [86]. This chapter investigates the effects of COVID-19 on EV users' charging behavior before, after, and during COVID-19 lockdown restrictions, using the same data from previous chapters. Data visualization using K-means and hierarchical clustering are analyzed. This work utilizes the users' connection and disconnection time to identify common charging pattern, where K-means clustering outperforms the hierarchical clustering for all three different scenarios modelled. In addition, prediction of collective charging session duration is achieved using ML Models, Random Forest and XgBoost. The results are evaluated using MAPE, achieving 14.6% and 15.1% for XgBoost and Random Forest respectively.[3]

4.2 Data Pre-processing

This chapter utilizes the same dataset from the previous chapter. However, the parameter consideration varies for this work. The important parameters from the dataset considered for this study includes the timestamps such as connection

[3]This chapter is a slightly modified version of our published paper, P. Rajagopalan and P. Ranganathan, "Electric vehicle charging behavior prediction using machine learning models," in 2022 IEEE Electrical Power and Energy Conference (EPEC), pp. 123–128, IEEE, 2022.

time, and load parameters such as kWh requested, kWh delivered recorded from the campus garage [66]. The study takes charging events from April 2018 to September 2021 into account, allowing the modelling of charging patterns before, during, and after the lockdown restrictions. The length of each charging session is determined by converting the connection and disconnection time as a suitable 24-hour time series scale.

In terms of data analysis, the goal is to figure out how the charging behavior changed prior to, after and during the period of the lockdown restrictions. The time frame of the lockdown period started when the World Health Organization (WHO) announced the outbreak of the global pandemic on March 11, 2020 [87]. To analyze the charging pattern and its shift, the number of charging sessions per month of three different years are visualized.

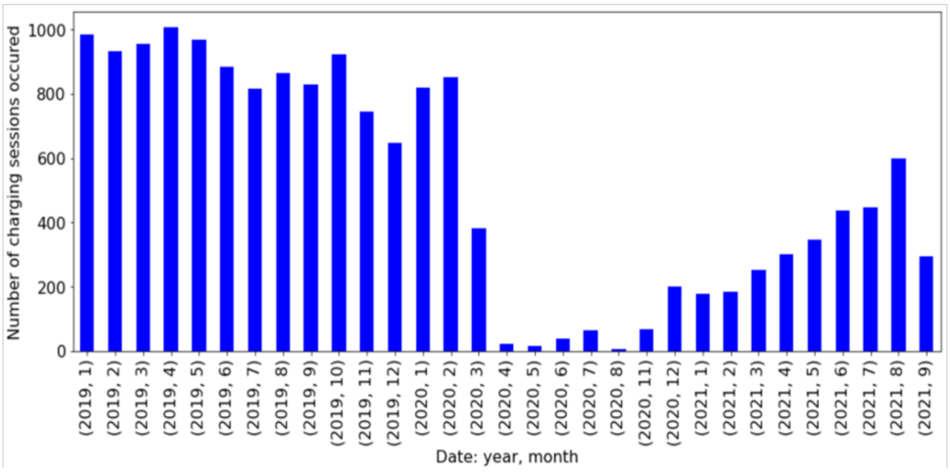


Figure 4.1: Charging Sessions Occurred Each Month from 2019-2021.

Figure 4.1 explains the significant decrease in charging activity between March and August 2020, there was nearly zero charging activity in terms of the number of hours during this period. Furthermore, the utilization of the charging station spiked in December 2020, due to the relaxation in lockdown restrictions. The analysis is further narrowed down to number of users per hour each year as shown in Figure 4.2.

This data visualization shows the peak hours of utilization of charging stations in 2019. We see that overnight charging preferred more when compared to

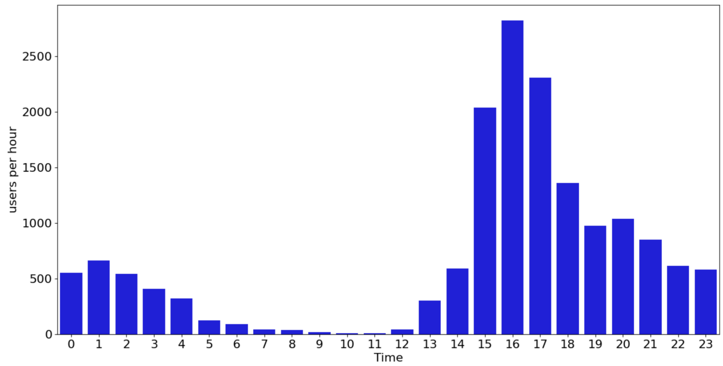
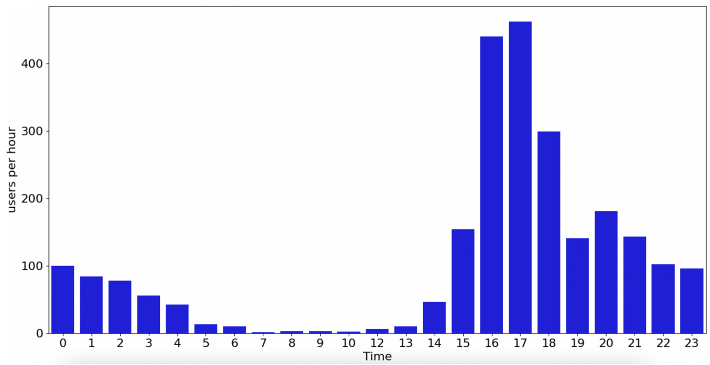
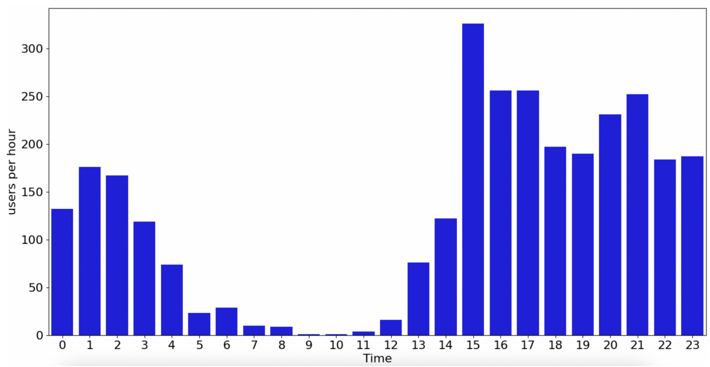


Figure 4.2: Number of EVs Arriving at Each Hour in 2019.

morning hours, and it reaches its peak by 4 p.m. With the help of this analysis, the number of users at different time frames are identified, thus contributing to the effective management of the charging infrastructure. Furthermore, to analyze the change in changing pattern in the following years, the same method is used to determine the number of users arriving at the charging station in the year 2020 and 2021 as represented in Figure 4.3.



(a) 2020.



(b) 2021.

Figure 4.3: Number of EVs Arriving at Each Hour During 2020-2021.

4.3 Clustering Techniques

In the clustering tasks, two popular clustering algorithms, K-means and hierarchical models are used to find common charging behavior. To determine the optimal number of clusters, before training the datasets, an elbow plot is followed. Figure 4.4 shows the elbow plot varying the number of clusters from 1 to 10 with the cluster Sum of Squared Error (SSE). As a result, the optimal number of clusters are set to three for both clustering algorithms.

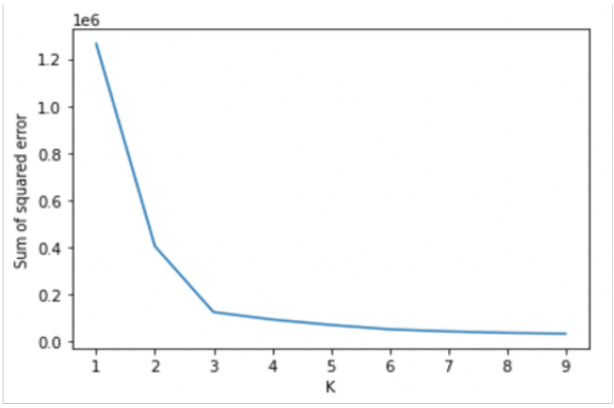


Figure 4.4: Elbow Plot for Deciding the Number of Clusters.

In K-means clustering, data points are assigned to k-centroid points at random. This method picks the right values for initial centroid, called K-means seeding. The number of clusters is denoted by the letter K. The data points are then assigned to new centroids in an iterative process based on their similarity [88]. The centroids are computed in the meantime, and each iteration is updated. The procedure is repeated until the algorithm converges and the cluster labels remain unchanged as the flow is illustrated in Figure 4.5. Each charging session from 2019, 2020, and 2021 is considered separately to implement in K-means and hierarchical clustering algorithms. To perform the clustering, the charging behaviors were identified by selecting the connection and disconnection times. The clustering was performed in Python using the Scikit-learn library [77].

In Figure 4.6, Cluster zero represents the overnight users utilizing the charging stations from 12AM Cluster one represents the overlap between the users in the evening hours. Silhouette index metrices are used to validate the quality of the

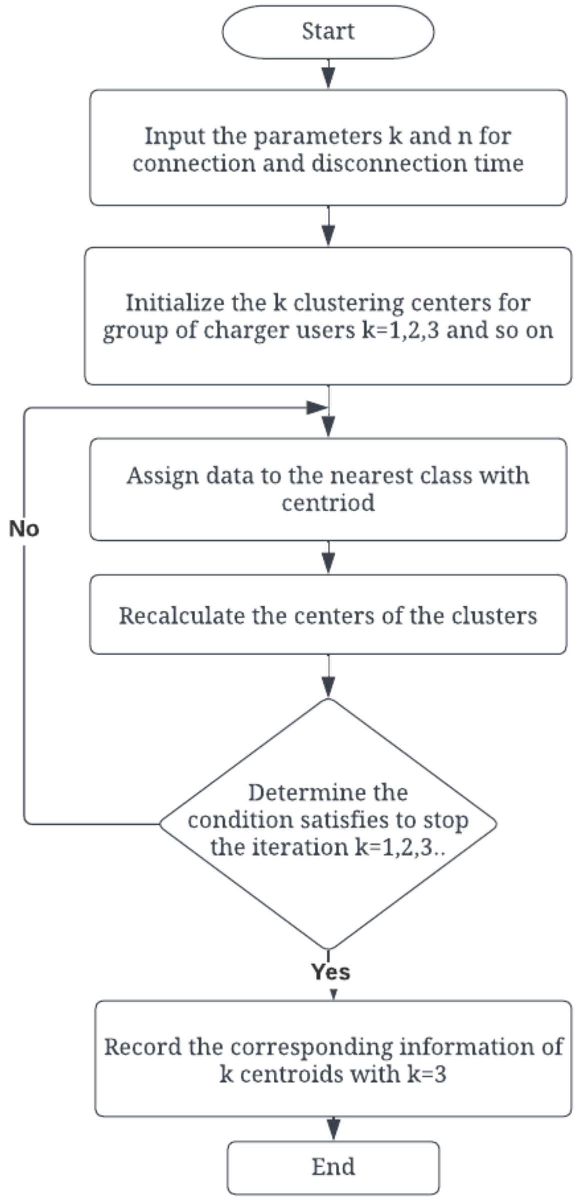
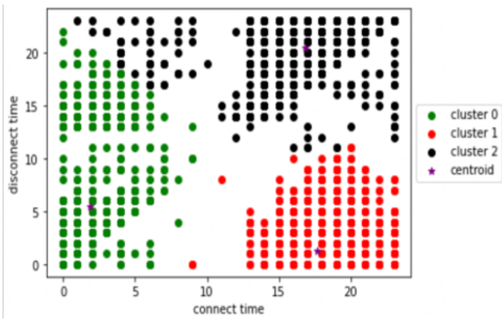


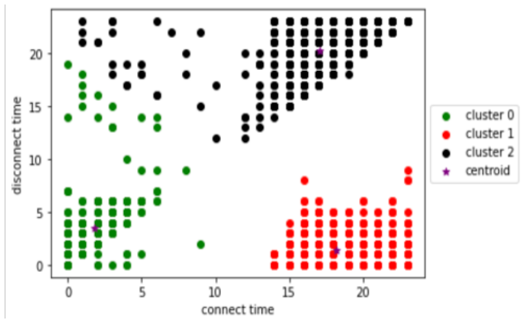
Figure 4.5: Flowchart for K-means Clustering Algorithm Based on Connection and Disconnection Time.

clusters. The results of this clustering technique are summarized in Table 4.1 silhouette indices evaluated for clustering methods.

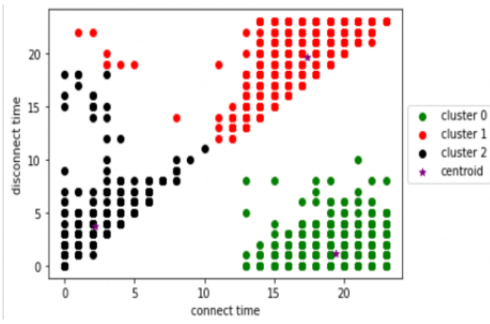
The silhouette index for the clusters formed using the K-means algorithm for the user connection and disconnection time were taken for 2019, 2020 and 2021. The Silhouette values are computed as 0.7726, 0.78799, and 0.7585 respectively. The hierarchical clustering algorithm produced similar number of clusters, as represented in Figure 4.7. By comparing the results, the K-means algorithm slightly



(a) 2019.



(b) 2020.



(c) 2021.

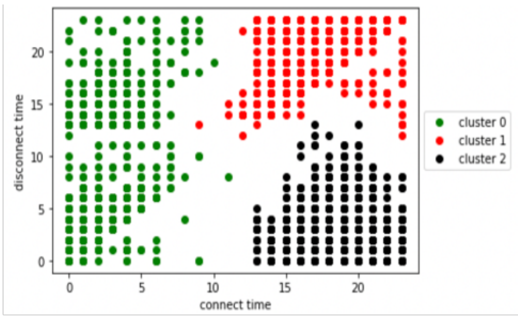
Figure 4.6: K-means Clustering Results for Connect and Disconnect Time for 2019-2021.

Table 4.1: Silhouette Indices for Clustering Algorithms.

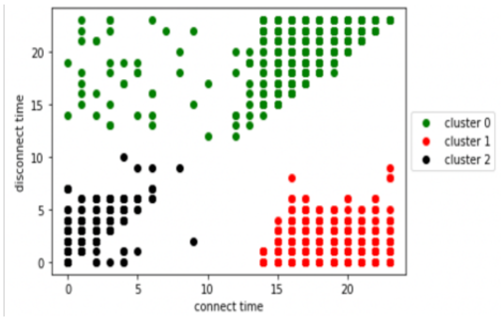
ML Models	**Silhouette Indices**		
	2019	***2020***	***2021***
K-Means	0.7726	0.7879	0.7585
Hierarchical	0.7542	0.7839	0.7518

outperformed the hierarchical clustering algorithm based on its accuracy metrics. The agglomerative hierarchical clustering algorithm for the user connection and disconnection time using Ward's method [89] is calculated with the internal validation scores which achieved 0.766, 0.783 and 0.758 respectively.

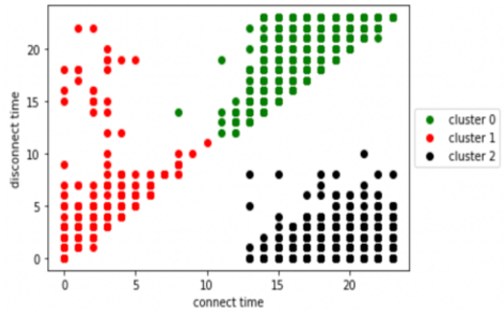
Although the formation and the number of clusters is similar between each year, the charging station utilization based on number of users varies by each cluster.



(a) 2019.



(b) 2020.



(c) 2021.

Figure 4.7: Hierarchical Clustering Results for Connection and Disconnection Time for 2019-2021.

4.4 Methodology

4.4.1 XgBoost Regression Model

XgBoost algorithm was developed by Dr.Chen in 2016, this model converts multiple weak learners into strong learners [90]. This approach can be regarded as an ensemble model of Classification and Regression Tree (CART). The predicted value corresponding to function space of CART are expressed as:

$$\hat{y}_{Xi} = \sum_{m=1}^{M} f_m(x_i), f_m \in F \tag{4.1}$$

where $\hat{y}_{Xi}$ is the predicted value of the ith sample, $f_m(x_i)$ is the predicted value of the ith sample in the mth tree, m represents the number of CART in the model.

Objective function is defined in Equation (4.2).

$$Obj = \sum_{i=1}^{\eta} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_m) \tag{4.2}$$

where η is the number of samples, l represents the second-order derivable loss function of the predicted value. $\Omega(f_m)$ is the regularization term.

The difference between connection and disconnection timestamps is used in the experimental analysis for predicting session duration with ML models such as RF and XgBoost with different hyperparameter settings as shown in Table 4.2.

Table 4.2: Hyperparameters for RF and XgBoost Algorithms.

ML Models	**Hyperparameters**	
	Parameter	**Value**
RF	n-estimators	10
	Random State	10
XgBoost	n-estimators	200
	Max Depth	1

4.5 Session Duration Prediction: Results and Discussion

For the ML models, the prediction errors of the session duration are calculated using MAPE metrics. Table 4.3 explains the comparative analysis of the predicted session duration of both ML models implemented. The proposed method reduces the prediction error.

$$MAPE \;=\; \frac{1}{n}\sum_{i=1}^{n} |\frac{y_i - y_p}{y_i}| \tag{4.3}$$

Table 4.3: Error Metrics - MAPE for Predicting Session Duration Using RF and XgBoost Models.

ML Models	**MAPE (Percentage)**	
	2018-2019	**2021**
RF	15.16	19.64
XgBoost	14.63	17.82

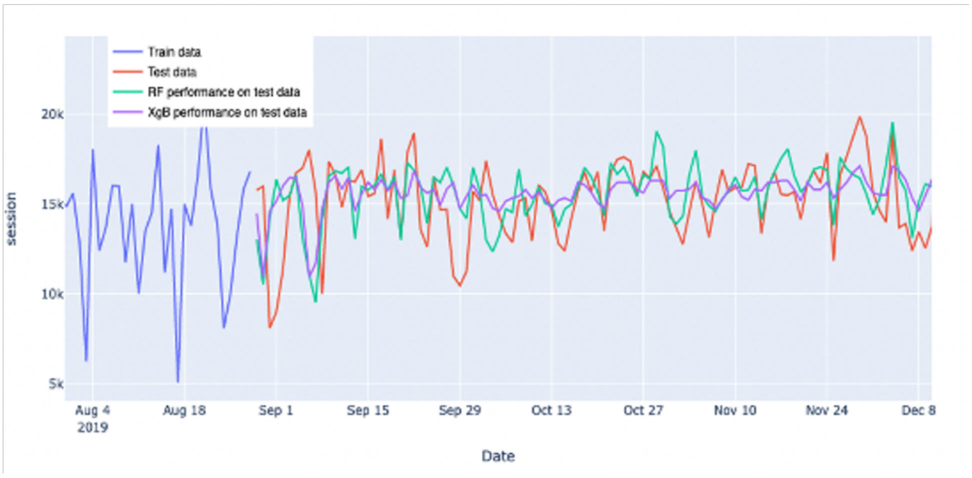


Figure 4.8: Performance of ML Models Over the Test Data of Charging Session Duration.

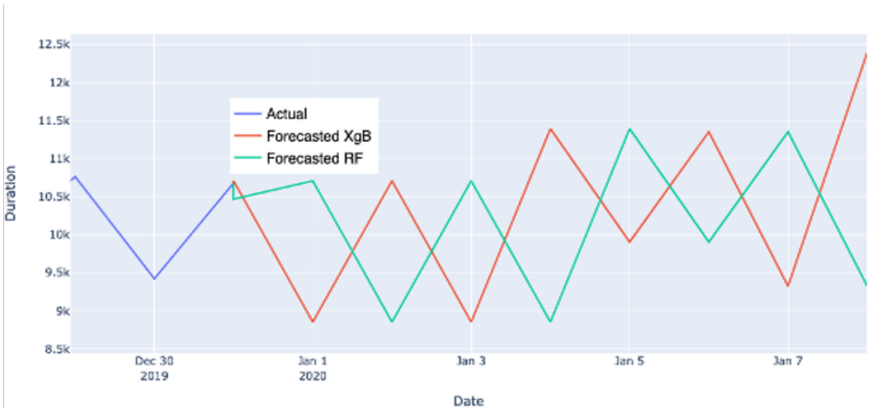


Figure 4.9: Charging Session Duration Forecast for 10-days Ahead.

Figure 4.8 represents the performance of the ML models over test data of the session duration from 2018-2019. The data is obtained from the difference in connection and disconnection hours. As the pandemic had the greatest impact on charging hours, the session duration considered is based on years from 2018 and 2019. The overall session duration is split into an 80%-20% ratio for the purpose of training and test validation of the ML models. Session duration that occurred from the period of April 2018 to September 2019 is considered as the value to train. Test values are considered from September to December 2019 for the ML models such as Random Forest [91] and XgBoost [92]. Figure 4.9 represents the forecast of the session duration for the next 10 days from the test value in the validation.

4.6 Conclusion

This chapter has utilized the unsupervised clustering techniques to examine the charging pattern before, after, and during lockdown restrictions. According to the preliminary results, there is a noticeable shift in time of utilization of charging stations, because of the pandemic lockdown measures. In terms of the session duration prediction results, XgBoost model performed better, considering the data before (14.63%) and after (17.82%) the COVID-19 data. There were no charging sessions recorded between the beginning of August and the middle of November 2020, but a significant increase noticed, starting from the middle of the year 2021. If this uncertainty in charging behavior continues, charging station usage will fluctuate on a regular basis, potentially affecting charging station planning and user convenience. In addition, in a economically fast growing state like California, the number of users is likely going to rise in the near future. As a result, this work is provided in order to identify a common pattern among users of charging stations, as well as to predicting charging duration, which is critical for charging station development.

Chapter 5

Conclusion and Future Work

An EV charging station designed to charge a fleet of EVs necessitates the focus of user convenience, specifically addressing the supply and demand imbalances. Thus, the use of ML models are proved to be valuable tools in load forecasting and estimating the number of users based on the idle time spent by the users and excess energy from the charging station. The main contribution of this work are as follows:

1. **Load Forecasting and Estimation of Number of Users based on Idle Time**: ML models such as SARIMA, RF and NN were utilized to perform short-term load forecasts with respect to connection time and kWh delivered from the EV charging station. The preliminary result indicates that all of the three models do agree that Monday's forecast yield a lower MAE values (SARIMA - 6.48, RF - 5.69, NN - 4.93), while Tuesdays, Thursdays, and weekends contain large errors. The chapter also investigated the estimation of number of users available for V2G service based on Idle time parameter. This is achieved using SVC and LR models. The results showed that historical Wednesdays have more number of users (1823). These can be useful for connecting back to the grid.

2. **Prediction of Excess Energy and Estimation of Number of Users based on Excess Energy**: The results of this study indicates the importance of application of ML models in capacity estimates for V2G services and prediction of excess energy from the charging station. This chapter compared four machine learning techniques, SVR, GBR, RF and LSTM, for predicting excess energy. From the preliminary results, it is observed that LSTM outperformed other models, with MAPE scores of 3.13 for Mondays, 4.21 for Tuesdays and

3.37 for weekends data respectively. This chapter also investigated the potential number of users likely to connect for V2G service based on 15% and 30% of excess kWh available from the charging stations using DT and KNN algorithms. Thus, V2G services are related to predicting excess kWh and estimation number of users as they rely on effective EV charging station management.

3. **Session Duration Prediction**: The pandemic has created a significant impact in both EV sales and the utilization of public charging infrastructure. Thus, the study of charging behavior before, after, and during lockdown restrictions plays a vital role in charging station management. This chapter has utilised the K-means and hierarchical clustering techniques to study the charging behavior. The models are evaluated using silhouette indices, based on the results, it is observed that K-means clustering technique performed better with 0.7726, 0.7879 and 0.7585 for the historical data 2019,2020 and 2021 respectively. In addition, the session duration from the charging station is predicted using RF and XgBoost, it is evident that XgBoost performed better considering the data before (14.63%) and after (17.82%) the COVID-19 data.

Therefore, ML plays a major role in studying the real time charging station located in California and in analyzing the capacity of users for V2G operations to adapt to the dynamic conditions.

- ML model enables the prediction of load and the excess energy from the charging station which helps to meet the demand efficiently.

- For real-time user convenience and adaption, ML models trained with historical behavior assists in understanding the user behavior. The charging sessions during the pandemic was identified distinct by their charging pattern.

- To efficiently plan and execute the V2G operation, ML algorithms supports in estimating the number of users based on the features such as idle time and excess energy.

With the rapid growth of EVs and it's significant contribution as distributed energy resources (DER), it is important to assess the user behavior, load management and the capacity of V2G services based on the aggregated users from the charging station. To investigate these parameters effectively, ML and DL models are more fast and reliable.

The limitations of this study includes the length of the dataset when it is further divided to weekdays and weekends. A more robust training model can be developed with the larger dataset. Future work could utilize streaming or dynamic dataset to study the excess energy available from the charging stations and can provide real-time data to EV aggregators.

Publication from this Thesis Work

- P. Rajagopalan, J. Thornby, and P. Ranganathan, “Short-term electric vehicle demand forecasts and vehicle-to-grid (V2G) idle-time estimation using machine learning,” in 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC), pp. 1279–1286, IEEE, 2023.

- P. Rajagopalan and P. Ranganathan, “Electric vehicle charging behavior prediction using machine learning models,” in 2022 IEEE Electrical Power and Energy Conference (EPEC), pp. 123–128, IEEE, 2022.

- P. Rajagopalan and P. Ranganathan, ”Predicting Excess Energy and Estimating Users for Vehicle-to-Grid (V2G) Services Using Machine Learning,” (Submitted to Ubiquitous Computing, Electronics and Communication Conference, 2023).

REFERENCES

[1] Iea, "Executive summary – global ev outlook 2022 – analysis." [Online]. Available: https://www.iea.org/reports/global-ev-outlook-2022/executive-summary

[2] "Energy systems and infrastructure analysis." [Online]. Available: https://www.anl.gov/esia/light-duty-electric-drive-vehicles-monthly-sales-updates

[3] "Alliance for automotive innovation." [Online]. Available: https://www.autosinnovate.org/posts/papers-reports/get-connected-2023-q1

[4] "Benefits from network of ev chargers across u.s." [Online]. Available: https://driveelectric.gov/files/2030-charging-network.pdf

[5] "Ev charging infrastructure trends in first quarter 2023." [Online]. Available: https://afdc.energy.gov/files/u/publication/electric_vehicle_charging_infrastructure_trends_first_quarter_2023.pdf

[6] "Internal revenue service report on ev purchase." [Online]. Available: https://www.irs.gov/credits-deductions/credits-for-new-clean-vehicles-purchased-in-2023-or-after

[7] L. Swagel, "Co2 in transportation sector," Dec 2022. [Online]. Available: https://www.cbo.gov/publication/58861

[8] D. Keith, "Challenges in ev adoption," Dec 2022. [Online]. Available: https://climate.mit.edu/explainers/electric-vehicles

[9] "Implementation of v2g strategies in california." [Online]. Available: https://docs.cpuc.ca.gov/PublishedDocs/Published/G000/M355/K104/355104591.PDF

[10] "V2g opportunities - report by american public power association." [Online]. Available: https://www.publicpower.org/periodical/article/public-power-utilities-others-pursue-vehicle-grid-opportunities

[11] D. Sun, Q. Ou, X. Yao, S. Gao, Z. Wang, W. Ma, and W. Li, "Integrated human-machine intelligence for ev charging prediction in 5g smart grid," *EURASIP Journal on Wireless Communications and Networking*, vol. 2020, no. 1, pp. 1–15, 2020.

[12] H. Moon, S. Y. Park, C. Jeong, and J. Lee, "Forecasting electricity demand of electric vehicles by analyzing consumers' charging patterns," *Transportation Research Part D: Transport and Environment*, vol. 62, pp. 64–79, 2018.

[13] R. Philipsen, T. Brell, W. Brost, T. Eickels, and M. Ziefle, "Running on empty–users' charging behavior of electric vehicles versus traditional refueling," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 59, pp. 475–492, 2018.

[14] Y. Tang, J. Zhong, and M. Bollen, "Aggregated optimal charging and vehicle-to-grid control for electric vehicles under large electric vehicle population," *IET Generation, Transmission & Distribution*, vol. 10, no. 8, pp. 2012–2018, 2016.

[15] Z. Wu, J. Hu, J. Wu, and X. Ai, "Optimal charging strategy of an electric vehicle aggregator in ancillary service market," in *2019 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia)*. IEEE, 2019, pp. 3223–3228.

[16] K. Singh and A. Singh, "Behavioural modelling for personal and societal benefits of v2g/v2h integration on ev adoption," *Applied Energy*, vol. 319, p. 119265, 2022.

[17] W. Wen, S. Yang, P. Zhou, and S. Gao, "Impacts of covid-19 on the electric vehicle industry: Evidence from china," *Renewable and Sustainable Energy Reviews*, vol. 144, p. 111024, 2021.

[18] S. Shahriar and A. Al-Ali, "Impacts of covid-19 on electric vehicle charging behavior: Data analytics, visualization, and clustering," *Applied System Innovation*, vol. 5, no. 1, p. 12, 2022.

[19] J. R. Helmus, M. H. Lees, and R. van den Hoed, "A data driven typology of electric vehicle user types and charging sessions," *Transportation Research Part C: Emerging Technologies*, vol. 115, p. 102637, 2020.

[20] S. Khan, B. Brandherm, and A. Swamy, "Electric vehicle user behavior prediction using learning-based approaches," in *2020 IEEE Electric Power and Energy Conference (EPEC)*. IEEE, 2020, pp. 1–5.

[21] P. Morrissey, P. Weldon, and M. O'Mahony, "Future standard and fast charging infrastructure planning: An analysis of electric vehicle charging behaviour," *Energy policy*, vol. 89, pp. 257–270, 2016.

[22] S. Viswanathan, J. Appel, L. Chang, I. V. Man, R. Saba, and A. Gamel, "Development of an assessment model for predicting public electric vehicle charging stations," *European Transport Research Review*, vol. 10, no. 2, pp. 1–15, 2018.

[23] W. Kanda and P. Kivimaa, "What opportunities could the covid-19 outbreak offer for sustainability transitions research on electricity and mobility?" *Energy Research & Social Science*, vol. 68, p. 101666, 2020.

[24] Y.-W. Chung, B. Khaki, T. Li, C. Chu, and R. Gadh, "Ensemble machine learning-based algorithm for electric vehicle user behavior prediction," *Applied Energy*, vol. 254, p. 113732, 2019.

[25] M. G. Flammini, G. Prettico, A. Julea, G. Fulli, A. Mazza, and G. Chicco, "Statistical characterisation of the real transaction data gathered from electric vehicle charging stations," *Electric Power Systems Research*, vol. 166, pp. 136–150, 2019.

[26] N. Sadeghianpourhamami, N. Refa, M. Strobbe, and C. Develder, "Quantitive analysis of electric vehicle flexibility: A data-driven approach," *International Journal of Electrical Power & Energy Systems*, vol. 95, pp. 451–462, 2018.

[27] R. Wolbertus, M. Kroesen, R. Van Den Hoed, and C. Chorus, "Fully charged: An empirical study into the factors that influence connection times at ev-charging stations," *Energy Policy*, vol. 123, pp. 1–7, 2018.

[28] T. Hu, K. Liu, and H. Ma, "Probabilistic electric vehicle charging demand forecast based on deep learning and machine theory of mind," in *2021 IEEE Transportation Electrification Conference & Expo (ITEC)*. IEEE, 2021, pp. 795–799.

[29] C. Liu, B. Sun, C. Zhang, and F. Li, "A hybrid prediction model for residential electricity consumption using holt-winters and extreme learning machine," *Applied energy*, vol. 275, p. 115383, 2020.

[30] 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, 2020.

[31] C. Scott, M. Ahsan, and A. Albarbar, "Machine learning based vehicle to grid strategy for improving the energy performance of public buildings," *Sustainability*, vol. 13, no. 7, p. 4003, 2021.

[32] X. Li, Y. Tan, X. Liu, Q. Liao, B. Sun, G. Cao, C. Li, X. Yang, and Z. Wang, "A cost-benefit analysis of v2g electric vehicles supporting peak shaving in shanghai," *Electric Power Systems Research*, vol. 179, p. 106058, 2020.

[33] M. Taljegard, L. Göransson, M. Odenberger, and F. Johnsson, "Electric vehicles as flexibility management strategy for the electricity system—a com-

parison between different regions of europe," *Energies*, vol. 12, no. 13, p. 2597, 2019.

[34] Y. Zhou, G. Yang, Y. Lei, M. Zhang, Y. Yang, H. Chen, H. Ye, and Q. Yu, "Ev load forecasting methods that consider the dispatchable potential in v2g system," in *2022 IEEE 5th International Conference on Electronics Technology (ICET)*. IEEE, 2022, pp. 303–307.

[35] C. O'Malley, L. Badesa, F. Teng, and G. Strbac, "Frequency response from aggregated v2g chargers with uncertain ev connections," *IEEE Transactions on Power Systems*, 2022.

[36] H. S. Nogay, "Estimating the aggregated available capacity for vehicle to grid services using deep learning and nonlinear autoregressive neural network," *Sustainable Energy, Grids and Networks*, vol. 29, p. 100590, 2022.

[37] F. Giordano, F. Arrigo, C. Diaz-Londono, F. Spertino, and F. Ruiz, "Forecast-based v2g aggregation model for day-ahead and real-time operations," in *2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. IEEE, 2020, pp. 1–5.

[38] A. Gautam, A. K. Verma, and M. Srivastava, "A novel algorithm for scheduling of electric vehicle using adaptive load forecasting with vehicle-to-grid integration," in *2019 8th International Conference on Power Systems (ICPS)*. IEEE, 2019, pp. 1–6.

[39] P. K. Mohanty, P. Jena, and N. P. Padhy, "Home electric vehicle charge scheduling using machine learning technique," in *2020 IEEE International Conference on Power Systems Technology (POWERCON)*. IEEE, 2020, pp. 1–5.

[40] D. Yu, M. P. Adhikari, A. Guiral, A. S. Fung, F. Mohammadi, and K. Raahemifar, "The impact of charging battery electric vehicles on the load profile in the presence of renewable energy," in *2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)*. IEEE, 2019, pp. 1–4.

[41] P. Rezaei and M. A. Golkar, "Economic load curve flattening by evs charge and discharge scheduling in the smart grid considering machine learning-based forecasted load," in *2021 11th Smart Grid Conference (SGC)*. IEEE, 2021, pp. 1–5.

[42] J. Cui, S. Liu, J. Yang, W. Ge, X. Zhou, and A. Wang, "A load combination prediction algorithm considering flexible charge and discharge of electric vehicles," in *2019 IEEE 10th International Symposium on Power Electronics for Distributed Generation Systems (PEDG)*. IEEE, 2019, pp. 711–716.

[43] M. Alamaniotis, "Synergism of deep neural network and elm for smart very-short-term load forecasting," in *2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*. IEEE, 2019, pp. 1–5.

[44] G. Van Kriekinge, C. De Cauwer, N. Sapountzoglou, T. Coosemans, and M. Messagie, "Day-ahead forecast of electric vehicle charging demand with deep neural networks," *World Electric Vehicle Journal*, vol. 12, no. 4, p. 178, 2021.

[45] Y. Shang, Z. Li, Z. Shao, and L. Jian, "Distributed v2g dispatching via lstm network within cloud-edge collaboration framework," in *2021 IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia)*. IEEE, 2021, pp. 1–8.

[46] M. Boulakhbar, M. Farag, K. Benabdelaziz, T. Kousksou, and M. Zazi, "A deep learning approach for prediction of electrical vehicle charging stations power demand in regulated electricity markets: The case of morocco," *Cleaner Energy Systems*, vol. 3, p. 100039, 2022.

[47] T.-Y. Ma and S. Faye, "Multistep electric vehicle charging station occupancy prediction using hybrid lstm neural networks," *Energy*, vol. 244, p. 123217, 2022.

[48] 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," *Applied sciences*, vol. 9, no. 9, p. 1723, 2019.

[49] Y. Lu, Y. Li, D. Xie, E. Wei, X. Bao, H. Chen, and X. Zhong, "The application of improved random forest algorithm on the prediction of electric vehicle charging load," *Energies*, vol. 11, no. 11, p. 3207, 2018.

[50] Q. Sun, J. Liu, X. Rong, M. Zhang, X. Song, Z. Bie, and Z. Ni, "Charging load forecasting of electric vehicle charging station based on support vector regression," in *2016 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*. IEEE, 2016, pp. 1777–1781.

[51] H. Wang, Y. Zhang, and H. Mao, "Load forecasting method of evs based on time charging probability," in *2018 International Conference on Power System Technology (POWERCON)*. IEEE, 2018, pp. 1731–1735.

[52] A. Gerossier, R. Girard, and G. Kariniotakis, "Modeling and forecasting electric vehicle consumption profiles," *Energies*, vol. 12, no. 7, p. 1341, 2019.

[53] B. Wang, B. Hu, C. Qiu, P. Chu, and R. Gadh, "Ev charging algorithm implementation with user price preference," in *2015 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. IEEE, 2015, pp. 1–5.

[54] Y. Bouhassani, N. Refa, R. Van Den Hoed *et al.*, "Pinpointing the smart charging potential for electric vehicles at public charging points," in *EVS32 symposium*, 2019.

[55] T. Mazhar, R. N. Asif, M. A. Malik, M. A. Nadeem, I. Haq, M. Iqbal, M. Kamran, and S. Ashraf, "Electric vehicle charging system in the smart grid using different machine learning methods," *Sustainability*, vol. 15, no. 3, p. 2603, 2023.

[56] D. Aguilar-Dominguez, J. Ejeh, S. F. Brown, and A. D. Dunbar, "Exploring the possibility to provide black start services by using vehicle-to-grid," *Energy Reports*, vol. 8, pp. 74–82, 2022.

[57] S. Kumar, N. Das, S. Deb, A. K. Goswami, and N. B. D. Choudhury, "Machine learning based demand supply system of an airport to minimize operational cost," in *2022 IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES)*. IEEE, 2022, pp. 1–6.

[58] S. Koohfar, W. Woldemariam, and A. Kumar, "Prediction of electric vehicles charging demand: A transformer-based deep learning approach," *Sustainability*, vol. 15, no. 3, p. 2105, 2023.

[59] S. Wang, C. Zhuge, C. Shao, P. Wang, X. Yang, and S. Wang, "Short-term electric vehicle charging demand prediction: A deep learning approach," *Applied Energy*, vol. 340, p. 121032, 2023.

[60] J. Shanmuganathan, A. A. Victoire, G. Balraj, and A. Victoire, "Deep learning lstm recurrent neural network model for prediction of electric vehicle charging demand," *Sustainability*, vol. 14, no. 16, p. 10207, 2022.

[61] X. Shen, H. Zhao, Y. Xiang, P. Lan, and J. Liu, "Short-term electric vehicles charging load forecasting based on deep learning in low-quality data environments," *Electric Power Systems Research*, vol. 212, p. 108247, 2022.

[62] Z. Yang, T. Hu, J. Zhu, W. Shang, Y. Guo, and A. Foley, "Hierarchical high-resolution load forecasting for electric vehicle charging: A deep learning approach," *IEEE Journal of Emerging and Selected Topics in Industrial Electronics*, vol. 4, no. 1, pp. 118–127, 2022.

[63] D. Zhou, Z. Guo, Y. Xie, Y. Hu, D. Jiang, Y. Feng, and D. Liu, "Using bayesian deep learning for electric vehicle charging station load forecasting," *Energies*, vol. 15, no. 17, p. 6195, 2022.

[64] M. P Sasidharan, S. Kinattingal, and S. Simon, "Comparative analysis of deep learning models for electric vehicle charging load forecasting," *Journal of The Institution of Engineers (India): Series B*, vol. 104, no. 1, pp. 105–113, 2023.

[65] M. D. Eddine and Y. Shen, "A deep learning based approach for predicting the demand of electric vehicle charge," *The Journal of Supercomputing*, vol. 78, no. 12, pp. 14 072–14 095, 2022.

[66] Z. J. Lee, T. Li, and S. H. Low, "Acn-data: Analysis and applications of an open ev charging dataset," in *Proceedings of the tenth ACM international conference on future energy systems*, 2019, pp. 139–149.

[67] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control.* John Wiley & Sons, 2015.

[68] S. I. Vagropoulos, G. Chouliaras, E. G. Kardakos, C. K. Simoglou, and A. G. Bakirtzis, "Comparison of sarimax, sarima, modified sarima and ann-based models for short-term pv generation forecasting," in *2016 IEEE international energy conference (ENERGYCON).* IEEE, 2016, pp. 1–6.

[69] [Online]. Available: https://www.statsmodels.org/dev/generated/statsmodels.tsa.stattools.adfuller.html#statsmodels.tsa.stattools.adfuller

[70] R. Mushtaq, "Augmented dickey fuller test," 2011.

[71] C. Li, Y. Tao, W. Ao, S. Yang, and Y. Bai, "Improving forecasting accuracy of daily enterprise electricity consumption using a random forest based on ensemble empirical mode decomposition," *Energy*, vol. 165, pp. 1220–1227, 2018.

[72] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "Liblinear: A library for large linear classification," *The Journal of Machine Learning Research*, vol. 9, pp. 1871–1874, 2008.

[73] "Sources of greenhouse gas emissions." [Online]. Available: https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions

[74] M. B. Arias and S. Bae, "Electric vehicle charging demand forecasting model based on big data technologies," *Applied energy*, vol. 183, pp. 327–339, 2016.

[75] B. Panda and R. K. Adhikari, "A method for classification of missing values using data mining techniques," in *2020 International Conference on Computer Science, Engineering and Applications (ICCSEA).* IEEE, 2020, pp. 1–5.

[76] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, pp. 273–297, 1995.

[77] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

[78] J. Che and J. Wang, "Short-term load forecasting using a kernel-based support vector regression combination model," *Applied energy*, vol. 132, pp. 602–609, 2014.

[79] Y. Zhang and A. Haghani, "A gradient boosting method to improve travel time prediction," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 308–324, 2015.

[80] J. H. Friedman, "Stochastic gradient boosting," *Computational statistics & data analysis*, vol. 38, no. 4, pp. 367–378, 2002.

[81] H. Shi, M. Xu, and R. Li, "Deep learning for household load forecasting—a novel pooling deep rnn," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 5271–5280, 2017.

[82] S. Kumar, L. Hussain, S. Banarjee, and M. Reza, "Energy load forecasting using deep learning approach-lstm and gru in spark cluster," in *2018 fifth international conference on Emerging Applications of Information Technology (EAIT)*. IEEE, 2018, pp. 1–4.

[83] W. Guo, L. Che, M. Shahidehpour, and X. Wan, "Machine-learning based methods in short-term load forecasting," *The Electricity Journal*, vol. 34, no. 1, p. 106884, 2021.

[84] G.-F. Fan, Y.-H. Guo, J.-M. Zheng, and W.-C. Hong, "Application of the weighted k-nearest neighbor algorithm for short-term load forecasting," *Energies*, vol. 12, no. 5, p. 916, 2019.

[85] S. Zhang, D. Cheng, Z. Deng, M. Zong, and X. Deng, "A novel knn algorithm with data-driven k parameter computation," *Pattern Recognition Letters*, vol. 109, pp. 44–54, 2018.

[86] A. Palomino, M. Parvania, and R. Zane, "Impact of covid-19 on mobility and electric vehicle charging load," in *2021 IEEE Power & Energy Society General Meeting (PESGM)*. IEEE, 2021, pp. 01–05.

[87] "World health organization (who)." [Online]. Available: https://covid19.who.int/

[88] T. M. Kodinariya, P. R. Makwana *et al.*, "Review on determining number of cluster in k-means clustering," *International Journal*, vol. 1, no. 6, pp. 90–95, 2013.

[89] F. Murtagh, "A survey of recent advances in hierarchical clustering algorithms," *The computer journal*, vol. 26, no. 4, pp. 354–359, 1983.

[90] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp. 1189–1232, 2001.

[91] M. R. Segal, "Machine learning benchmarks and random forest regression," 2004.

[92] B. Pan, "Application of xgboost algorithm in hourly pm2. 5 concentration prediction," in *IOP conference series: earth and environmental science*, vol. 113. IOP publishing, 2018, p. 012127.

APPENDICES

A Appendix A: GitHub Repository with Codes

Below is the GitHub repository link containing the code.

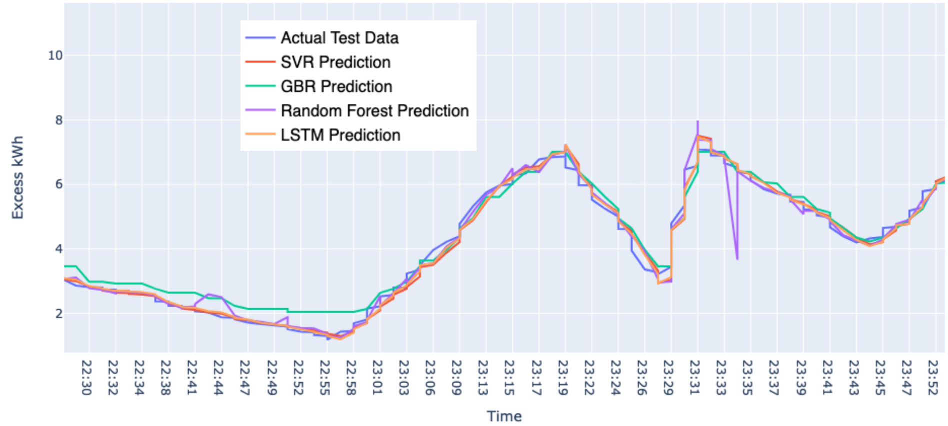
https://github.com/Prashanthrajagopal91/Prashanth_Thesis.git

B Appendix B: Visualization for predicting excess energy from charging station

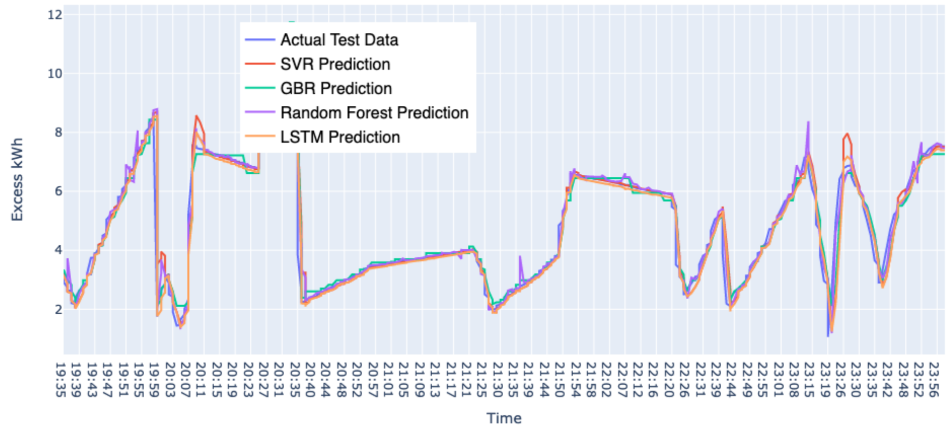
Below is the detailed experimental visualization for predicting excess energy from charging station using SVR, GBR, RF and LSTM.

C Appendix C: Visualization for prediction and forecasting of load parameters from charging station

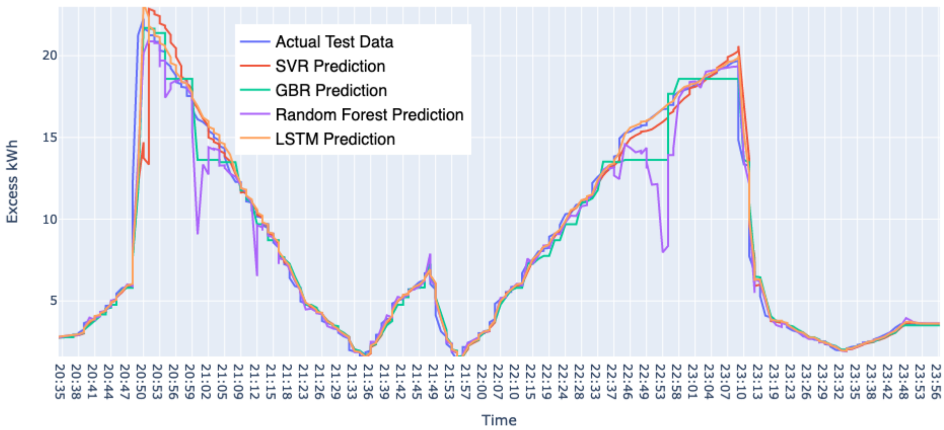
Below is the detailed experimental visualization for prediction and forecasting of kWh delievered from charging station using RF, NN and SARIMA models.



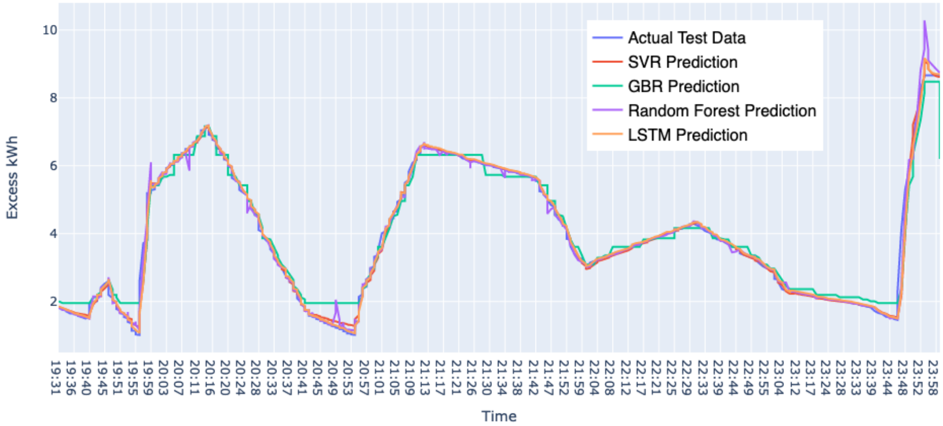
(a) Monday.



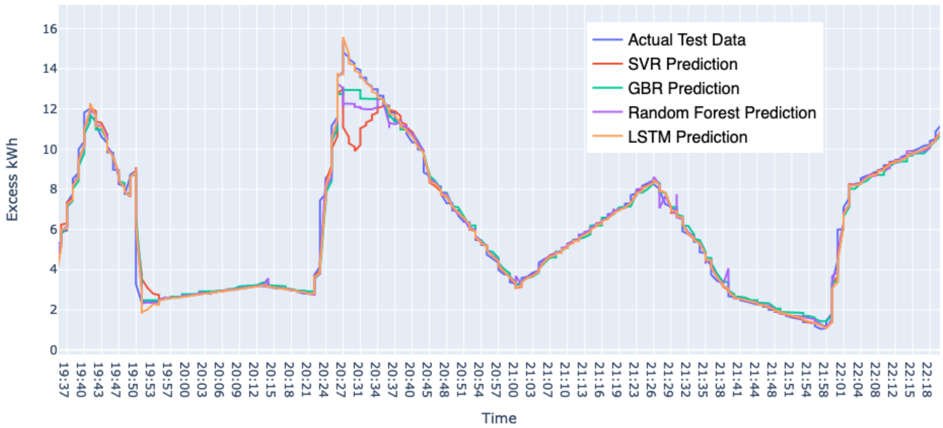
(b) Tuesday.



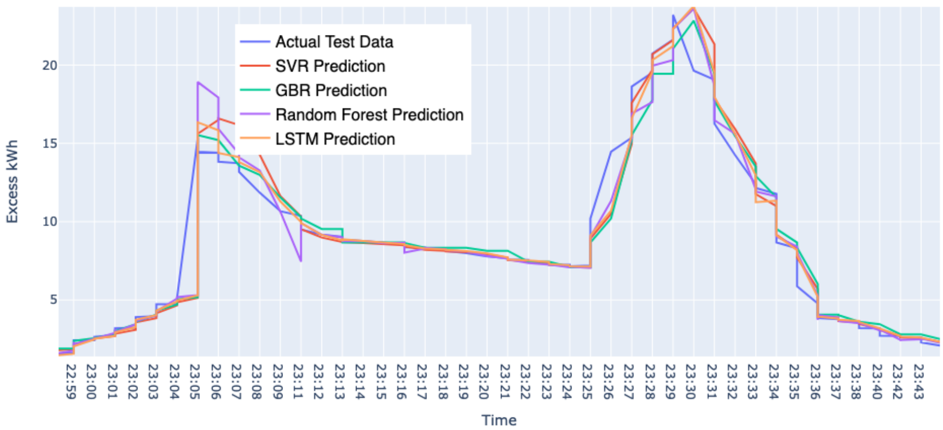
(c) Wednesday.



(d) Thursday.

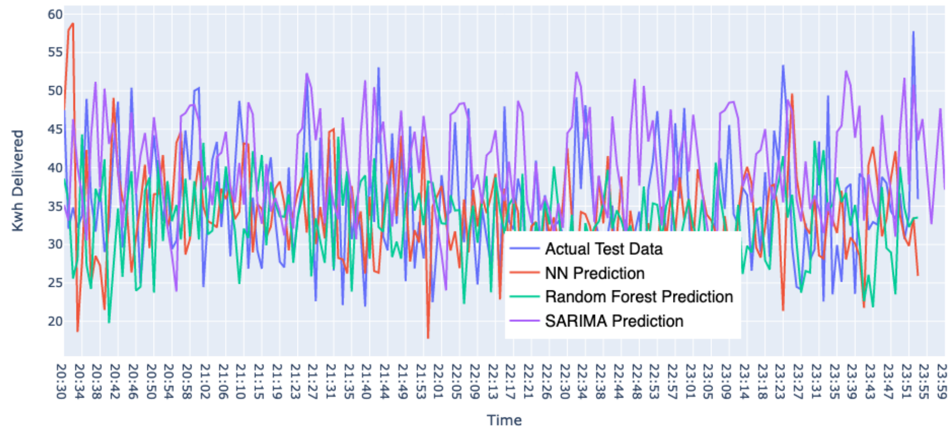


(e) Friday.

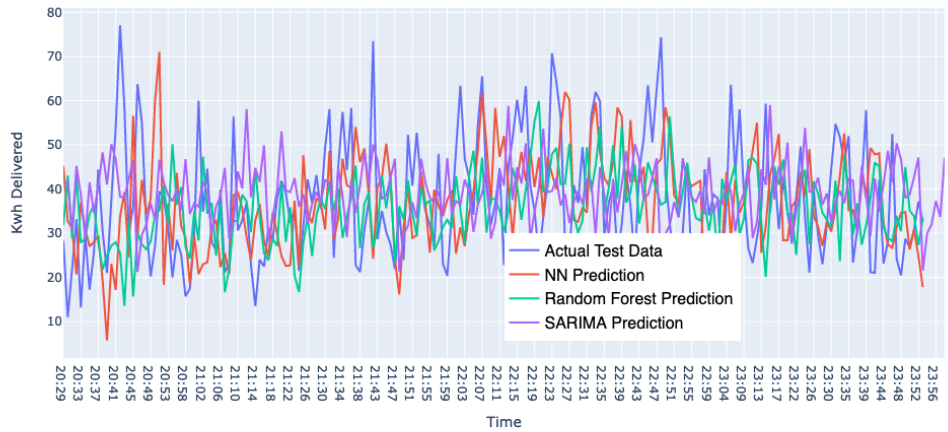


(f) Weekends.

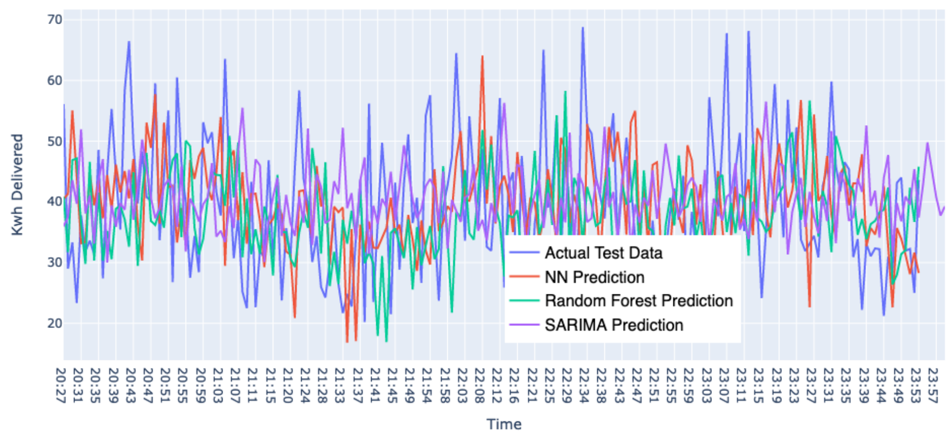
Figure B.1: Prediction of Excess kWh for Each Days within the Data Period of 2018-2021.



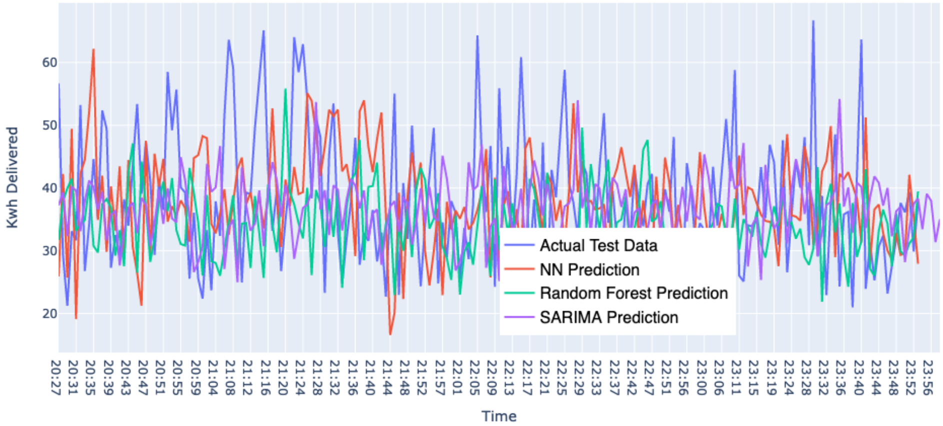
(a) Monday.



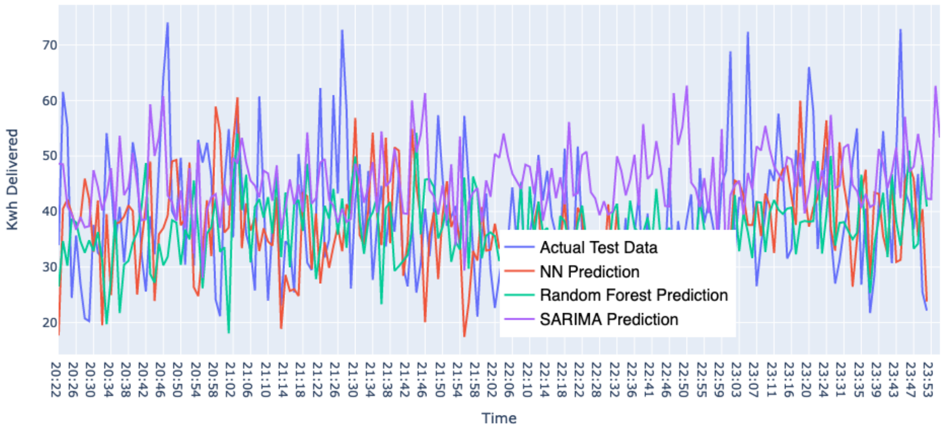
(b) Tuesday.



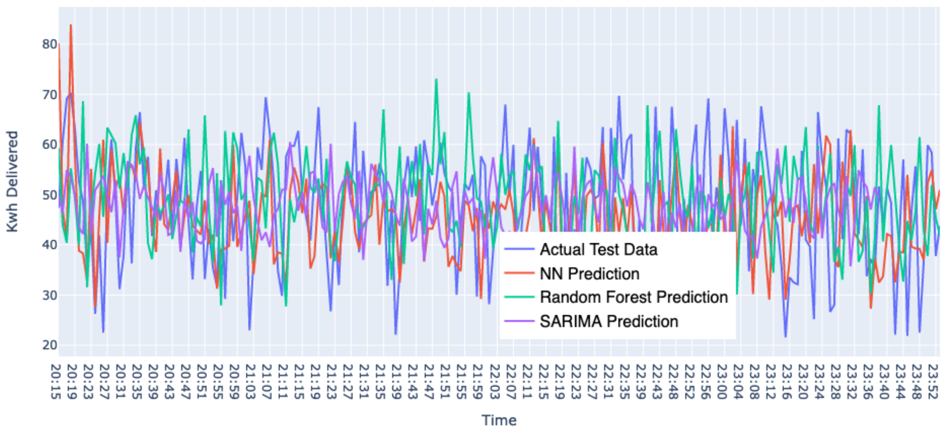
(c) Wednesday.



(d) Thursday.

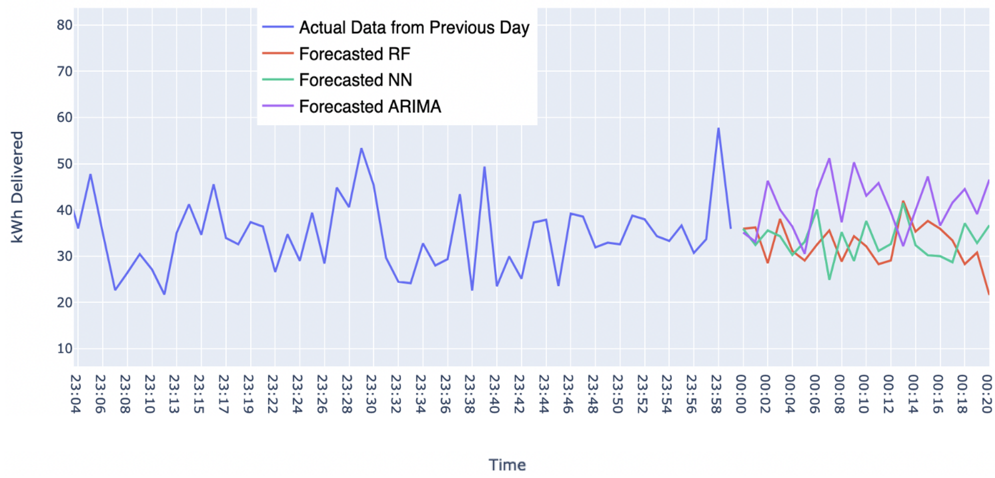


(e) Friday.

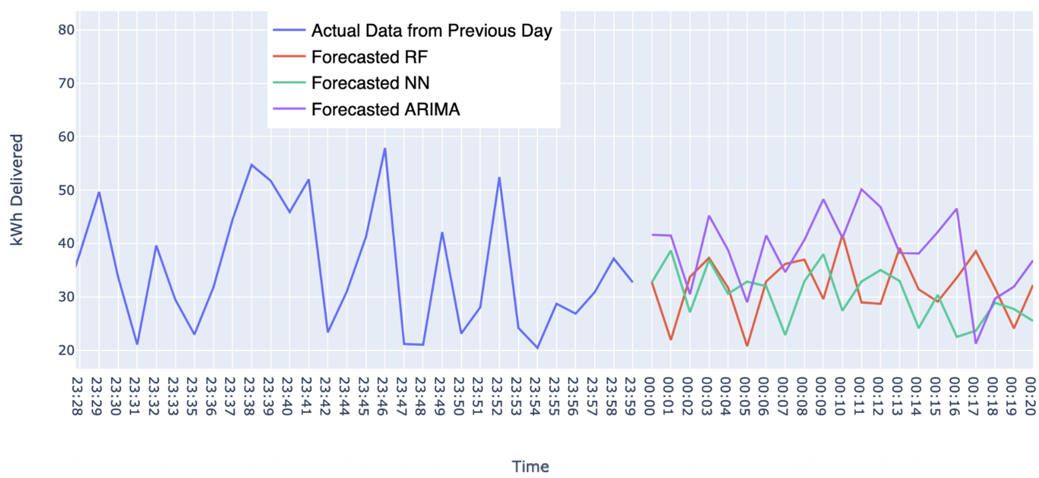


(f) Weekends.

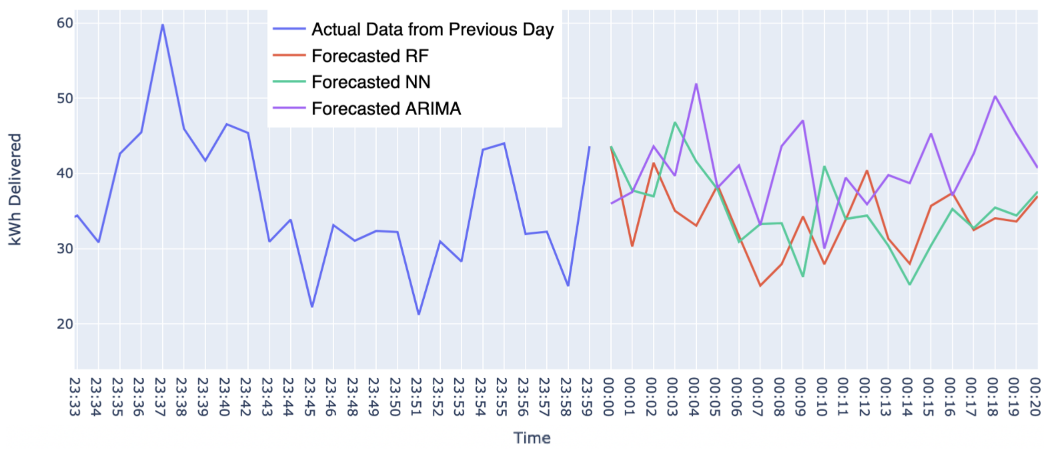
Figure C.1: Prediction of kWh Delivered using Different ML Models within the Data Period of 2018-2021.



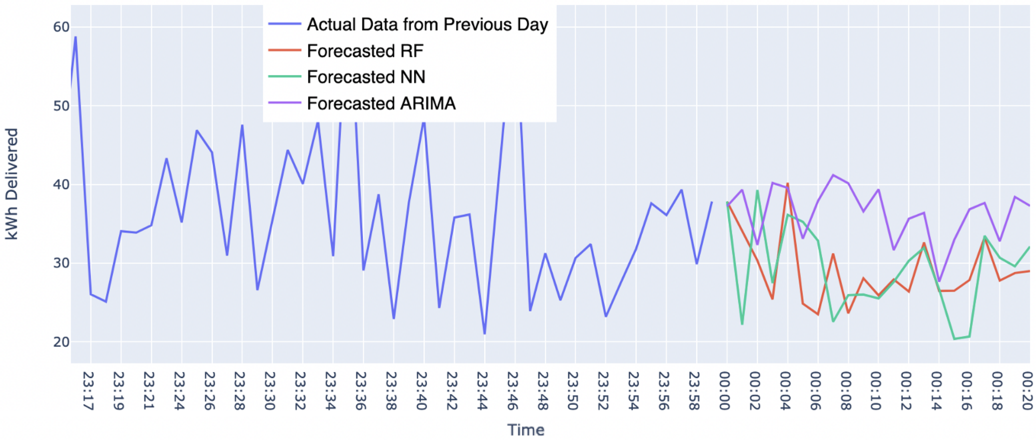
(a) Monday.



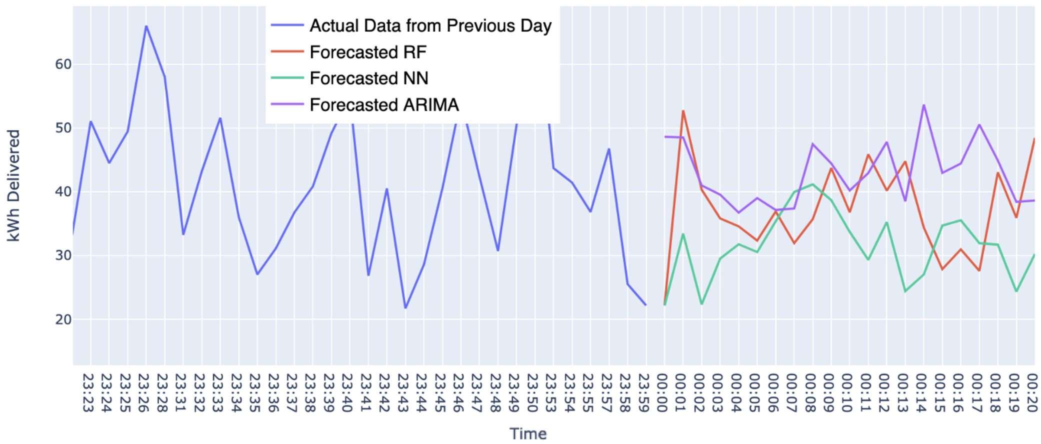
(b) Tuesday.



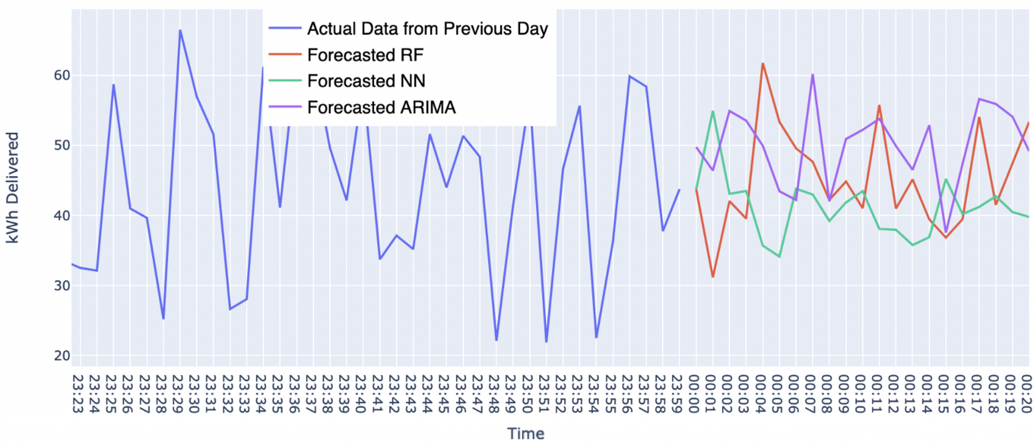
(c) Wednesday.



(d) Thursday.

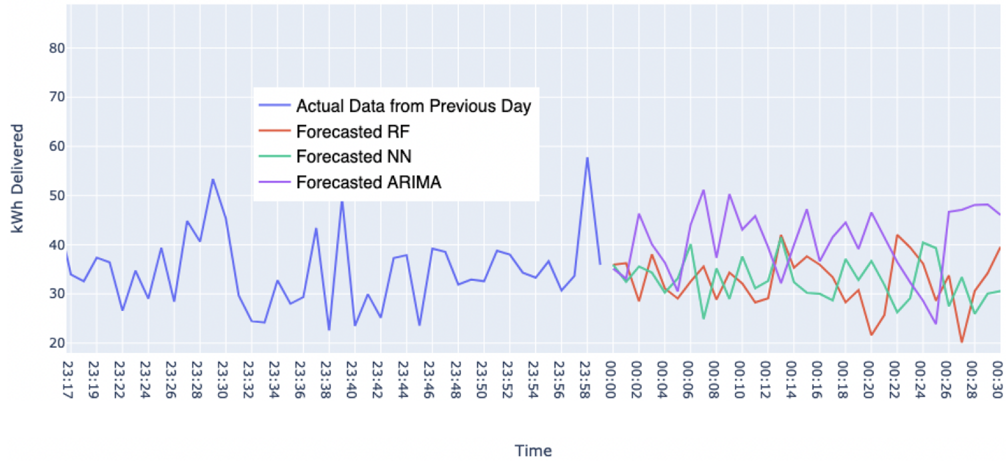


(e) Friday.

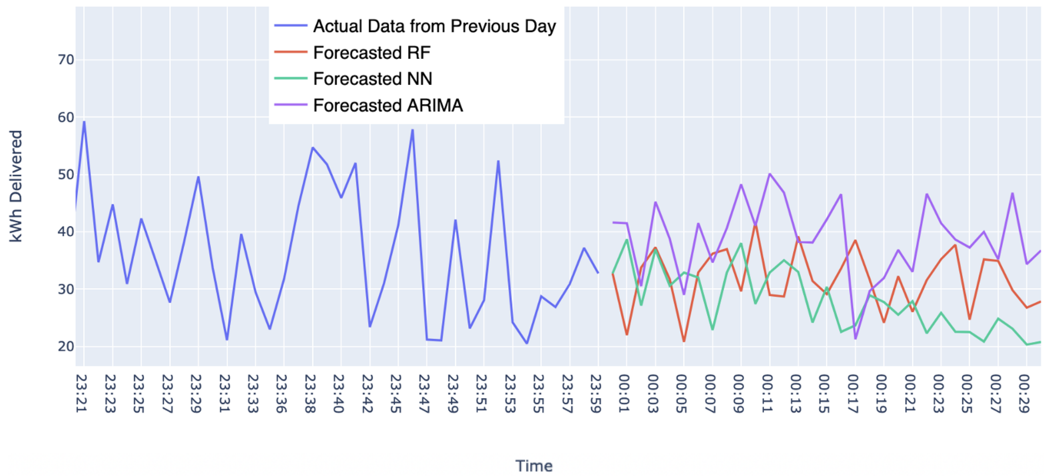


(f) Weekends.

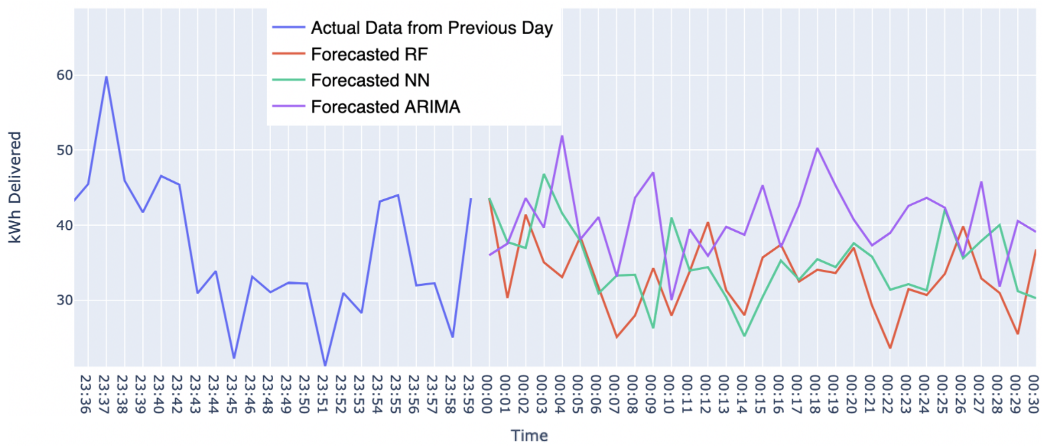
Figure C.2: 20-Minutes Ahead Forecast for kWh Delivered using Different ML Models.



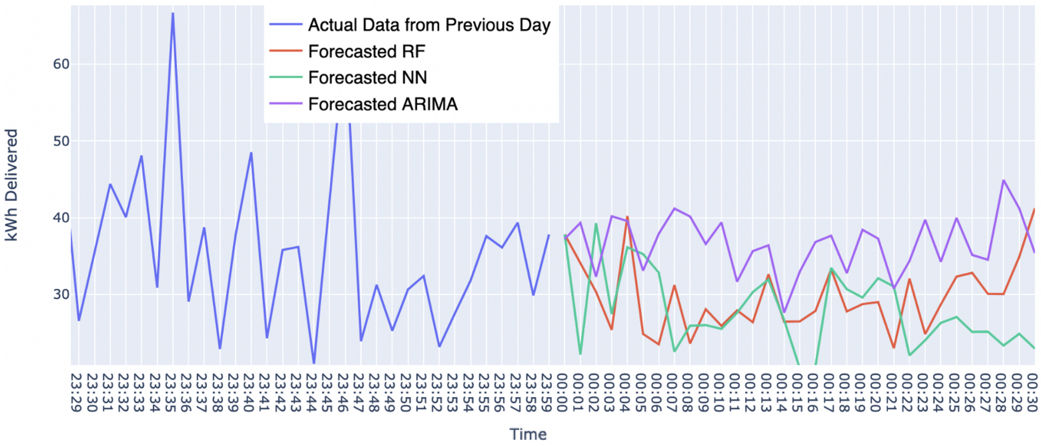
(a) Monday.



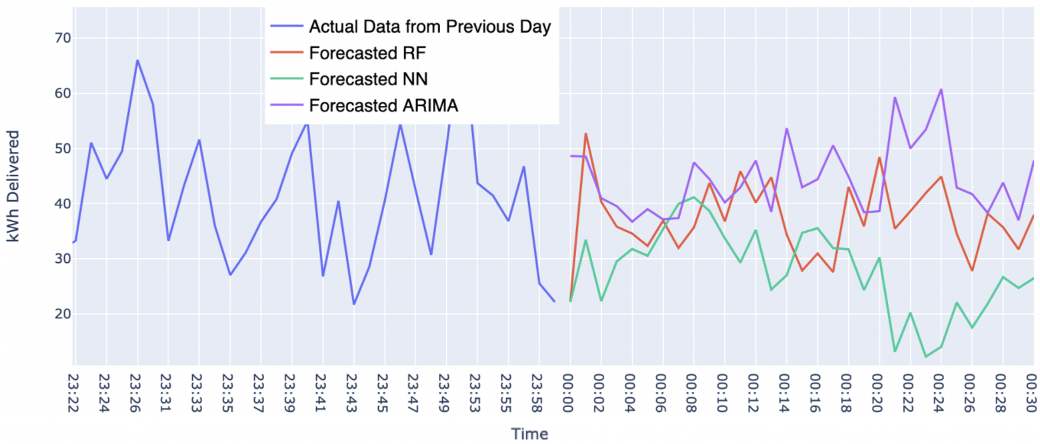
(b) Tuesday.



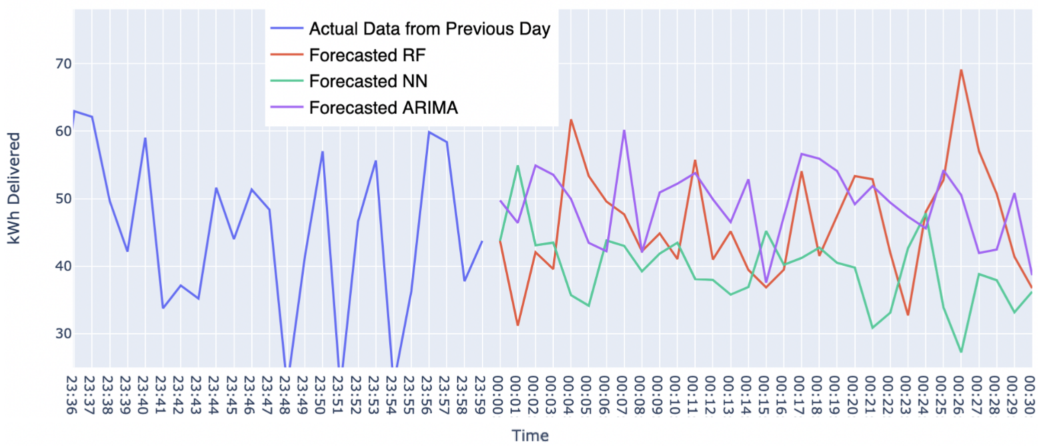
(c) Wednesday.



(d) Thursday.



(e) Friday.



(f) Weekends.

Figure C.3: 30-Minutes Ahead Forecast for kWh Delivered using Different ML Models.