Norwegian University
of Life Sciences

Master's Thesis 2021 30 ECTS
Faculty of Science and Technology

An analysis of electric vehicle user-behaviour at a smart charging station

Andreas Sandvik Høimyr
M.Sc. Environmental Physics and Renewable Energy

Acknowledgements

This thesis marks the end of my studies here at the Norwegian University of Life sciences. I have really enjoyed my tenure here at Ås and I have made both good friends and memories that I will cherish.

I would like to thank my supervisor Heidi S. Nygård for her guidance and continuous feedback on my work. Our weekly meetings have really helped me through this semester.

I would also like to thank Bendik Nybakk Torsæter and Rubi Rana at SINTEF Energy Research for providing me with the inspiration this thesis builds on and their feedback on my work. I really appreciated our conversations and your insightful inputs on my work.

Ås, May 2021

Andreas sandvik Høimyr

Abstract

Due to the electrification of the transport sector, it has become a concern for grid operators, about the challenges the increasing energy demand from the transport sector could impose on the power grid. Hence, to enable the further electrification of the transport sector, and avoid heavy additional investments in power grid infrastructure, there has become a demand of a further understanding of electric vehicle (EV) user-behaviour. Thus, the goal of this thesis was to analyse how EV owners charged their EV, and how this varied between EV owners charging at a charging station. This was done with a data-driven approach, using a dataset recorded at a real charging station, the Adaptive Charging Network (ACN) at the Institute of Technology in California, USA. Furthermore, it was attempted to build Machine Learning models to predict the charging session duration, and the energy delivered in a charging session, using the dataset. Another part of the goal of this thesis was to present some factors that could be useful to record at a charging station, for analysing EV user-behaviour, and present an overview of some open source datasets that could be used for analysing aspects of EV user-behaviour.

In the analysis it was observed that more EV owners charged at the ACN on weekdays, than on weekends. EV owners charging on weekdays followed a charging pattern similar to that expected for a workplace charging station. Furthermore, the EV owners charging at the ACN on weekdays left their EV connected for longer than the EV owners charging on weekends. Moreover, observations suggested that EV owners who charged on weekdays tended to leave their EV connected for longer than the duration the EV was charging. When dividing the EV owners into distinct charging groups, based on how often the EV owners charged their EV, it was found that the EV owners charging the most frequently tended to leave their EV connected for longer than the other EV owners. The machine learning resulted in models with a limited performance, with an R2-score of around 0.50 for both the model predicting charging session duration and the model predicting energy delivered in a charging session. Lastly, based on the observations in the analysis, factors that could be useful to record for analysing EV user-behaviour were proposed. These factors were the time of connection, time of disconnection, an id to separate between EV owners, and the price of charging.

Sammendrag

Elektrifiseringen av transportsektoren har ført til at kraftsystemoperatører har uttrykt bekymringer knyttet til utfordringer det økende energibehovet fra transportsektoren kan påføre kraftnettet. Derfor, for å muliggjøre en videre elektrifisering av transportsektoren, og å unngå tunge tilleggsinvesteringer i kraftnettinfrastruktur, har det oppstått et behov for en ytterligere forståelse av brukeratferd av elektriske kjøretøy. Målet med denne oppgaven var å analysere hvordan elbileiere ladet elbilen sin, og hvordan brukeradferden til elbileiere variert mellom elbileiere som ladet på en ladestasjon. Dette ble utført med en datadrevet analyse av et datasett fra en reel ladestasjon, Adaptive Charging Network (ACN), ved Institute of Technology i California, USA. I tillegg, ble det forsøkt å bygge maskinlæringsmodeller for å predikere varighet på ladeøkt, og energien som leveres i løpet av en ladeøkt, med informasjonen i datasettet. Det var også et mål for denne oppgaven å presentere en oversikt over noen offentlig tilgjengelige datasett, som kan brukes til å analysere aspekter ved elbil brukeratferd, og å presentere noen faktorer som kan være nyttige å registrere ved en ladestasjon, for å analysere elbil brukeradferd.

I analysen ble det blant annet observert at flere elbileiere ladet på ACN på hverdager, enn i helgene. Videre, ble det observert at elbileierne som ladet på hverdager, fulgte et lademønster som liknet på det som forventes for en ladestasjon tilknyttet en arbeidsplass. I tillegg ble det observert at elbileierne som ladet på ACN på hverdager, var tilkoblet ladestasjonen lenger enn elbileierne som ladet i helgene. Det ble også observert at elbileierne som ladet på hverdager hadde en tendens til å la elbilen være tilkoblet vesentlig lengere enn det som tilsynelatende var nødvenidg for å lade elbilen. Når elbileierne ble delt inn i forskjellige ladegrupper, basert på hvor ofte de ladet elbilen, ble det observert at elbileierne som ladet oftest, hadde en tendens til å la elbilen være tilkoblet lengere enn de andre elbileierne. Maskinlæringen resulterte i modeller med en begrenset ytelse, med en R2-verdi på rundt 0,50 for både modellen som predikerte varighet på ladeøkt, og modellen som predikerte energi levert i ladeøkt. Til slutt, basert på observasjonene i analysen, ble faktorer som kunne være nyttige å registrere for å analysere brukeratferd foreslått. Disse faktorene var tilkoblingstid, frakoblingstid, en id som kan skille mellom elbileiere og prisen for lading.

Contents

List of Figures

List of Tables

Abbreviations

A.C.	Alternating Current
ACN	Adaptive Charging Network
API	Application Programming Interface
ASA	Adaptive Scheduling Algorithm
D.C.	Direct current
DLC	Direct Load Control
DSO	Distribution System Operator
DT	Decision tree
EV	Electric Vehicle
EVSE	Electric Vehicle supply equipment
ICEV	Internal Combustion Engine Vehicle
ILC	Indirect Load Control
KNN	KNearest Neighbors
MAE	Mean Absolute error
OLS	Ordinary Least Squares
R2	Coefficient of determination
RF	Random Forest
SSE	Sum of Squared Error
SST	Sum of Total Squares
STD	Standard Deviation
TSO	Transmission system operator

1. Introduction

1.1 Motivation

In 2015, Norway along with 195 other countries, adopted the Paris agreement. The treaty set a goal to limit global warming to 2° Celsius [45]. To achieve this long-term temperature goal, countries would need to cut their climate gas emissions. To comply with the Paris agreement the Norwegian government, in cooperation with the EU, committed to cut their climate gas emissions by 40 % by 2030 [13], [27]. Norway also became the third country in the world to report reinforced climate goals targeting a reduction in climate gas emissions of 50% to 55% by 2030, and 90% by 2050, to become a low-emission society [27].

To achieve the climate goals set by the Norwegian government, climate gas emissions across all sectors must be cut, including the transport sector. According to Statistisk Sentralbyrå (SSB), the transport sector in Norway were responsible for almost a third of all carbon emissions in 2017. Of this, 56% came from road traffic, making it the second biggest contributor to climate gas emissions in Norway, after the oil and gas sector [42]. A path to reduce emissions from the transport sector is to transition from Internal Combustion Engine Vehicles (ICEV) to emission free vehicles. Thus, to reduce emissions from the transport sector, the Norwegian government implemented several benefits for buying and owning EVs to help transition the car park from ICEV to EVs, until EVs becomes a viable option. According to the Norwegian Ministry of Transport and Communications, these benefits included no VAT and no large one-time fee when buying an EV [29]. As a result of this, the Norwegian EV market has greatly increased and by the end of 2020 there were more than 340 000 registered EVs in Norway. This was an increase of 370 % since 2015 [43].

Initially, the EV owner population consisted of commuters who either charged at home or at work, and the need for public charging stations was limited. But, as the number of EV owners increased, and due to a desire to further increase the EV population it became a substantial need for a widespread public charging infrastructure. To accommodate this, the Norwegian government implemented several support schemes to build and expand an extensive charging infrastructure [24]. In 2021 there were more than 3000 charging stations in Norway with more than 18 000 charging points [33].

This rapid and extensive electrification of the car park has led to a large and growing power demand from the transport sector. This power demand comes on top of the ordinary residential and commercial load and has created a concern from both Distribution System Operators (DSO) and Transmission System Operators (TSO) , about the large-scale integration of EVs in the distribution grid. For example, in January 2021 there were set multiple demand peaks records on an hourly basis in Norway [3]. This led to a request from grid operators that EV owners should try to avoid charging in peak hours of the day [31].

It is feared that a widespread implementation of EVs could potentially have an impact on the electricity distribution infrastructure. The power grid can handle a high penetration of EVs. However, several studies have found that, if either the power for EV charging increases or a large amount of EVs in the same area charge at the same time, it could create challenges for some components in the distribution grid and harm the power quality. Especially in areas with low capacity and in congested grid situations. E.g. a study by DNV GL and Poyry Management Consulting on behalf of NVE [10], and a study by S. Johansson et al. [19].

Due to the concern of how uncontrolled EV charging can impact the power grid, the topics of load balancing and peak flattening through smart charging schemes have become popular topics of research. Several studies have found that the implementation of smart charging could help the power grid handle a higher penetration of EVs and avoid heavy investments in new infrastructure. E.g. a study by DNV GL and Poyry Management Consulting on behalf of NVE [10], a study by S. Johansson et al. [19], and by Z. J. Lee et al. [23]. However, to incorporate reliable smart charging schemes that both fulfil the power systems technical constrains and maximises the service provided to EV owners, it is necessary to have a good understanding of EV user-behaviour.

Traditionally, EV user-behaviour has been modelled as similar to that of ICEVs, or modelled with statistical and mathematical approaches. In reality EV charging behaviour is different than ICEV user-behaviour, and a lot of the unpredictability and individuality of EV user-behaviour is difficult to capture in a mathematical model [39]. Thus, it is vital for further research and improvement of smart charging methods to have vast data of EV charging available. According to S. Shahriar et al. [39], with the emergence of Big data analytics and machine learning in recent years, solving the EV charging problem, and modelling EV user-behaviour has been seen as a suitable application of machine learning. Machine learning models can be trained on historical charging data, and data about weather and traffic, to accurately capture trends in charging behaviour. Moreover, they can produce accurate predictions, that can be used independently, or with other algorithms, for smart scheduling .

1.2 Objective of thesis

The electrification of the transport sector is set to proceed due to both political goals, and EVs becoming favoured options when people buy new vehicles. Therefore, it has become of interest to create better models of EV charging demand, and produce smart charging schemes for EV charging, to avoid heavy additional investment in grid infrastructure. Hence, there has become a broad focus on gaining a better understanding of EV user-behaviour at charging stations.

The goal of this thesis, was to analyse how EV owners charged their EV at a public charging station, and identify factors that could have impacted how EV owners charged their EV, using a data-driven approach. This was done by analysing certain aspects of EV user-behaviour observed at a charging station, and analysing how these aspects varied between the EV owners, using data recorded at a real charging station. The aspects of EV user-behaviour that were investigated in this thesis were the time of Connection, the duration EV owners left their EV connected, and the amount of energy EV owners received in a charging session. To analyse the differences in EV user-behaviour between EV owners, the EV owners were categorised into charging groups based on how frequently they charged at the ACN. These charging groups where then used to analyse differences in connection time, duration of charging session, and energy delivered in charging sessions, for the EV owners in each charging group. Furthermore, machine learning was used to predict aspects of EV user-behaviour displayed at a charging station. Namely, the charging session duration and the energy delivered in a charging session. With the observations from the data-driven analysis and the machine learning, this thesis also aimed to propose some factors that could be advantageous to record at charging stations, for analysing EV user-behaviour.

It was desired to conduct the analysis using real data, recording real EV user-behaviour displayed at public charging stations. However, when searching for data to use, it became clear that there was a limited selection of open-source data. Furthermore, there was a lack of a general overview of applicable datasets. Hence, an important aspect of this thesis, became to identify, and create an overview of some publicly available datasets, that could be used to analyse EV user-behaviour. Moreover, it was also decided to conduct the analysis using one of the found open-source datasets. Even though, the motivation for this thesis was mainly due to the electrification of the Norwegian transport sector, a dataset recorded at a charging station at the California Institute of Technology (Caltech) was chosen for the analysis, the ACN-dataset [22]. The main reason for this was due to the limited selection of available data. However, the ACN-dataset also provided a lot of information, and overall enabled a detailed analysis of EV user-behaviour. Additionally, the ACN-dataset was recorded at a smart charging station, which allowed for a novel investigation on how smart scheduling affected the EV owners who charged at the charging station. This investigation was mainly focused on analysing if EV owners received their requested energy before their requested departure time, and analysing discrepancies between the the requests of the EV owners and the service provided by the charging station.

2. Theory

2.1 Electric vehicle charging stations

An EV charging station refers to the system used to charge an EV. There are three methods to charge an EV, inductive, conductive, or by changing the battery. Conductive charging is currently the preferred type of EV charging due to its superior efficiency and lower cost [8]. Conductive charging stations can be classified into two types of charging systems, on-board and off-board systems. An on-board charging system is inside the EV. It allows for flexible charging as the owner can charge their EV at a grounded Alternating current (A.C.) power socket, and the on-board system rectifies the A.C. signal to a Direct Current (D.C.) signal when charging the vehicle. The charging power for on-board charging is limited due to the systems weight and size. Off-board charging systems on the other hand, are built at fixed locations and are not limited by these factors. With off-board charging A.C. power from the grid is rectified locally and charges the EVs battery directly. This allows for higher charging levels, fully charging the EVs battery faster [14].

EV charging stations are also categorised based on the power level the charging system can provide. There are mainly three types, Level 1, Level 2, and D.C. charging stations [14]. Figure 2.1 shows a simple visualisation of the general structures of these three types of charging stations.

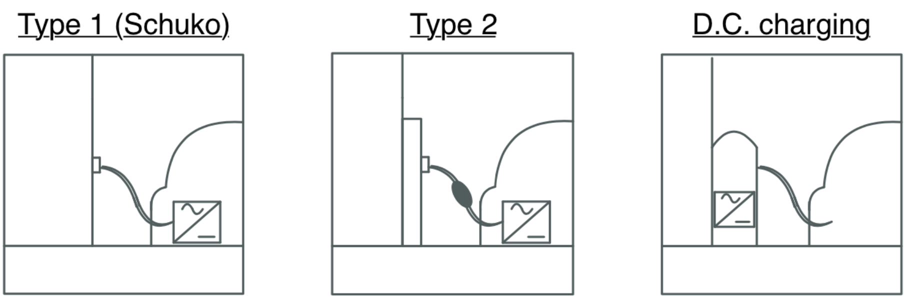


Figure 2.1: A simple visualisation of the structure of a type 1, a type 2 and a D.C. charging station. The rectifier represents the position of the charging system, either off-board or on-board.

Level 1 charging allows for charging an EV at a Shucko socket, a common grounded A.C. socket [8]. Level 2 charging utilises a dedicated A.C. Electric Vehicle supply equipment (EVSE) and allows for charging an EV with either 7 kW single phase power, or 22 kW three phase power, known as FastAC charging. For both Level 1 and Level 2 charging the EV is dependent of an on-board charging system that rectifies the A.C. signal to charge the battery [8]. The third type of charging station is the D.C. charging station. D.C. charging utilises an off-board D.C. EVSE which allows for higher power to charge the EV. The power supplied from a D.C. charger ranges all from the power level of Level 2 chargers and up to high power levels [8]. E.g. there are D.C. charging stations that allow for charging with 350 kW [18]. A charging station is referred to as a semi-fast or fast charging station if it allows for charging at 22 kW or more [33].

It is important to note that, at a charging station, there can be multiple outlets available to charge EVs. These outlets are often referred to as charging points. Hence, at a charging station, multiple EVs can charge simultaneously through different charging points.

2.2 Charging stations and charging patterns

The user-behaviour of EV owners can vary depending on the accessibility of a charging station. EV charging stations can usually be separated into three categories based on the access they provide for EV owners. Namely, private, public, and workplace charging stations. Private, or residential, charging stations provide charging for the owner of the charging station. Workplace charging stations are owned by a company and provides charging for the company's employees. Public charging stations are either commercially or privately owned, but provide charging for all EV owners who desire to charge.

Figure 2.2 shows expected charging patterns on weekdays for private, public, and workplace charging stations, as adapted from data available through Elaad.nl open data dashboard [11]. The Figure shows an expected distribution of connecting EVs during a day, based on real charging activity at charging stations in the Netherlands [11]. Figure 2.2 shows that, at private charging stations, most EVs connect to charge in the evening and at night. Moreover, it shows an expected peak of connecting EVs between 17 and 19, with many EVs connecting even later. For public charging stations this changes. Public charging stations tend to be used during the day, with the highest number of EVs connecting in the morning between 7 and 10, and in the evening between 17 and 19. Lastly, at workplace charging stations EVs mainly connect to charge during work hours. Figure 2.2 shows a distinct peak in connections in the morning, between 7 and 10, when employees arrive at work.

Referencing the distributions in Figure 2.2, an EV charging station can constitute in a

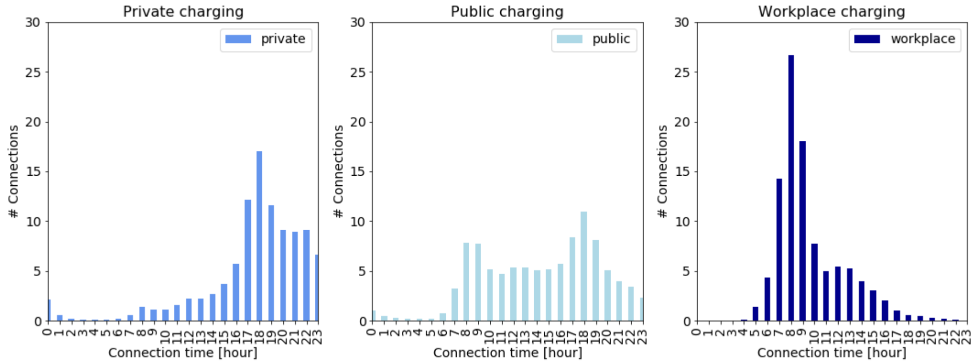


Figure 2.2: Expected distribution of when EVs connect to Private, public and workplace charging stations. Adapted from data available at Elaad.nl open data dashboard [11].

large power demand in peak hours, especially when charging is done at higher power levels. This can unfortunately result in challenges for the power grid, if it coincides with periods were the power grid is congested [19].

2.3 Power grid structure and operation

2.3.1 Power grid structure

Figure 2.3 displays a simple model of a power system, where A.C. power is generated at a power plant, injected into the power grid and transferred to the end-users through the transmission and distribution lines.

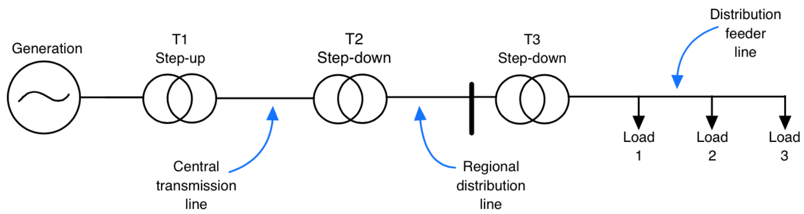


Figure 2.3: A simple presentation of a power system with one source of generation and one line transporting the generated power to three end-users.

In Figure 2.3, 3-phase A.C. power is generated at a power plant. The generated power is injected into the power grid at a fixed voltage [47]. This voltage is transformed in several transformers (T1, T2 and T3) before it reaches the three end-users at load 1, load 2, and load 3. In T1 the voltage is transformed up to a magnitude used in central transmission lines. As central transmission lines are usually long, power is transported at a high voltage [47]. E.g. In Norway the voltage in central transmission lines is usually in the region of 300 kV or 420 kV [35]. When the power is transferred to the area where it will be used, the voltage is transformed down. In T2 the voltage is transformed down to a magnitude used in regional distribution lines. E.g. in Norway the voltage in regional distribution lines is 22 kV or less [35]. Before the power is consumed by end-users, it is transformed down in the step-down transformer T3, and distributed through distribution feeder lines. The voltage in distribution feeder lines is usually 400 V or 230 V in Norway.

2.3.2 EV charging stations in the power grid

With respect to EV charging stations, they are usually connected in the low voltage, distribution feeder lines, like Load 1, Load 2, and Load 3 in Figure 2.3. For Level 1 charging, an EV is simply charged through a regular Schuko socket, at for example a residential home. Level 2 charging is done with a designated A.C. EVSE, and can be installed at home, or in a public location. D.C. charging uses a designated D.C. EVSE, and are usually found in public locations. There is however, a lot of research on the impact and opportunity of connecting fast charging stations, using a very high power to charge EVs, directly to the medium voltage, regional distribution grid instead. E.g. an article by S. Srdic and S. Lukic., investigated the challenges and opportunities of connecting extreme fast charging stations directly to the medium voltage grid [41].

2.3.3 Balancing generation and consumption

The main challenge for operating a power system is energy conservation [47]. For consumers to receive the power they request, an equal amount of power must always be generated. Whereas, power plants can regulate the generation, and consumers can choose when to use power, the transmission lines in the power grid can only transfer power, and do not have the option of storage [47]. Hence, it becomes the responsibility of grid operators to always balance the consumption and generation of power, in a power system, to ensure that end-users receive a sufficient power, at all times. To balance power generation and consumption, grid operators both forecast the power consumption in advance, and monitor it in real-time, making sure that the generation of power and consumption of power is always balanced [47].

2.4 Effects on the power grid

There are several probable grid impacts of EV charging. Widespread EV charging can result in a relatively large aggregated load demand that could contribute to an increased peak demand and increased system losses [10], [25]. Moreover, according to an article by C. H. Dharmakeerthi et al [9], the EV charging load is difficult to forecast as it depends on several factors that cannot simply be predicted in advance. E.g. the location, charging duration, connection time, and power consumption of individual EVs. Hence, widespread EV charging could contribute to violation of local and regional grid constrains, such as voltage limit violation and harmonic distortion.

2.4.1 Voltage limit violation

Voltage drops in long distribution feeders can be quite significant, and are dependent on the power demand [47]. Initially, generators inject power at a fixed voltage magnitude, which translate through several transformers to the fixed supply voltage for customers. As the power consumption increases, the current in the power grid increases and there is a drop in voltage according to Ohms law. System operators allow for some variation of voltage, as it is practically impossible to maintain a flat voltage profile. In Norway, operators allow a tolerance of ± 10 % of the nominal voltage for slow changes, and ± 5 % for rapid fluctuations [28].

Figure 2.4 shows an example, where voltage limits are violated in a distribution feeder.

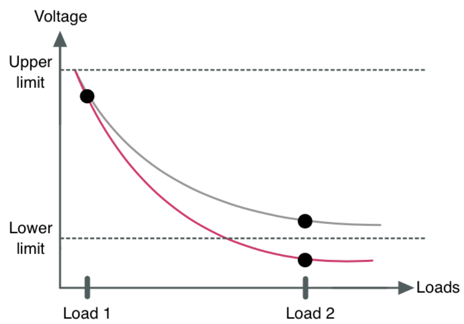


Figure 2.4: Voltage drop in a distribution feeder. A large load is connected in load 1, creating a significant voltage drop, such that load 2 receives a voltage lower than the allowed limit.

If a large inductive load, like many EVs charging simultaneously, connects to a distribution feeder in an already congested grid situation, the voltage may decrease beyond the allowed tolerance, as seen in Figure 2.4. This can lead to nuances such as dimming lights, but also more substantial consequences for sensitive commercial and industrial appliances, for the consumers connected to the distribution feeder. In the U.S., such voltage sags are responsible for an estimated 5 billion dollars in economic losses each year [47].

2.4.2 Harmonic distortion

A clean waveform refers to when oscillations of the A.C. signal follows the mathematical sine wave. Deviations from this sinusoidal wave is known as harmonic distortion.

Distortion to the voltage waveform is often caused by generators, while distortion to the current waveform is usually caused by loads [47]. According to A. Lucas et al. [25], current distortion is very common when it comes to nonlinear loads, such as EV charging systems. EV charging systems (On-board and off board charging systems) use power electronics switches to rectify A.C. power from the grid to D.C. power, to charge the EV battery. This process introduces distortion currents to the distribution grid which can contribute to distort the power signal [25]. The consequence of the distortion currents is that transformers are forced to operate at lower efficiencies, and they can damage the lifetime of the transformers due to overheating [47].

2.4.3 Loss of power and thermal heating of distribution lines

Another aspect of the distribution grid that can be affected by EV charging is the degradation of distribution lines through thermal heating. With EV charging being associated higher power demand peaks [16], it contributes to a higher current in distribution lines. A higher current (I) in the distribution lines leads to a higher loss of power (P) to thermal heating (Q) according to the I^2R-factor (equation 2.1) [49].

$$Q = P = I^2 R \tag{2.1}$$

Heating of lines may lead to damage and a shorted lifetime. As a line is heated up, it stretches from thermal expansion and sags. If it sags to far, the distortion of the line becomes irreversible, and it must be replaced [47].

2.5 State of the art: Smart charging

The problems associated with EV charging, such as increased peak demand, increased system losses, and power quality issues, are mainly due to uncontrolled EV charging. As mentioned, most problems only arise in congested grid situation if many EVs connect to the grid. For high power charging, these problems do however, become more prevalent. Nevertheless, according to a study by Z. J Lee et al. [23], it could be possible to make the power grid handle a higher penetration of EVs, and avoid heavy additional investment in grid infrastructure by incorporating smart charging for EV charging [23]. The goal of smart charging schemes is to maximise the service provided to EV owners at charging stations and fulfil the power systems technical constraints [40]. With smart charging it could be possible to move EV charging demand away from peak hours (peak shaving) and use the EV charging station as a flexible resource (load balancing) [20]. Smart charging can be categorised into two main types, Direct Load Control (DLC) , and Indirect Load Control (ILC) .

2.5.1 Direct Load Control

DLC approaches consists of methods were the charging station operator takes full control of the EVs charging regiment. They are usually done at an aggregated level to optimise the benefit of the system or station operator [20]. DLC can be used to move some of the EV charging load away from peak load hours. An example of a DLC approach is presented in an article by E. C. Kara et al. [20]. In the article it is assumed that an EV charging session begins when an EV connects to a charging point, and that EVs usually are done charging before they disconnect. The proposed approach was designed to move some of the charging load to this slack period before the vehicle disconnected [20]. The desired effect of this was to shave demand peaks by moving some of the charging load away from peak demand hours.

Another example of a DLC scheme is the ACN [23]. The ACN applies and adaptive charging algorithm (ASA) to schedule the charging session for each connecting EV. When an EV connects, its owner gives inputs on requested energy and desired departure time. The algorithm uses these inputs to optimally schedule the charging session, to prevent the ACN from breaking grid constraints, while delivering the requested energy before the requested departure. As more users connect, the ACN adapts to the new power demand. The ASA then adapts the scheduled charging of each connected vehicle to not breach grid constraints. A full break down and analysis of how the ASA works is given in an article by Z. J. Lee et al., 'Adaptive charging networks: A framework for smart electric vehicle charging [23].

2.5.2 Indirect Load Control

In ILC approaches, station operators do not directly control the charging regime, however, they try to influence the charging behaviour of EV owners with external factors such as dynamic pricing schemes [20]. An example of an ILC approach is the Deadline Differentiated Pricing presented by E. Bitar and Y. Xu [2]. The general idea of deadline differentiated pricing is to offer the customer different prices for charging based on their requested deadline of delivery. The longer an EV owner is able to delay their deadline, or departure, the cheaper the price of charging becomes. This will give the station operator flexibility to meet load demand and avoid breaking grid constrains [2]. The desired effect of the dynamic pricing scheme is that more users are willing to charge for longer or avoid charging in peak demand hours.

2.6 Machine learning

Machine learning is a subfield of Artificial intelligence involving self-learning algorithms that derive knowledge from data in order to make predictions. Machine learning can be divided into three main types, supervised learning, unsupervised learning and reinforcement learning [37]. Each of the three types are applied for different scenarios. For the understanding of this thesis only a description of supervised learning is included, along with a novel description of some popular supervised learning regression algorithms.

2.6.1 Supervised Learning

In supervised learning, algorithms are used to train a model by learning rules in input data with known output signals (target variables). Figure 2.5 displays the general concept of a supervised learning algorithm.

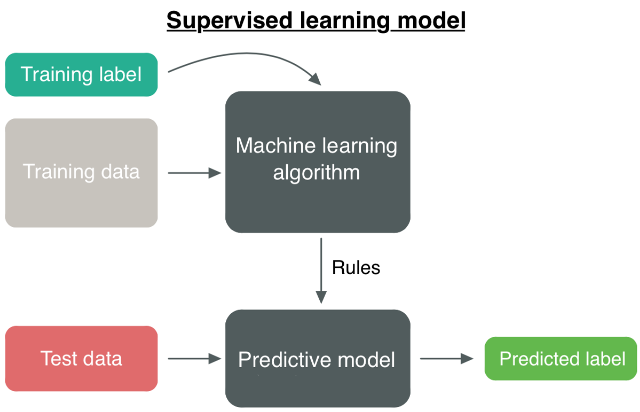


Figure 2.5: The general concept of a supervised learning algorithm as adapted from [37]. Training data and training label are input the algorithm. The algorithm learns the rules and combination of weights which most accurately models the relationship between the input data and input training label. The learned rules and combinations are then used when the algorithm is input new data to predict its target variable (predicted label)

Figure 2.5 shows how the supervised learning model takes a training dataset consisting of m rows of data (samples), each row consist of n columns of information (features), and a training label consisting m values representing the output signals (target variables) of the samples. The model learns the relationship between the features that give the corresponding target variable for each training sample. This learning process is known as fitting the model and it produces the predictive model. The predictive model can then be given new data and predict its target variables based on the observations in the learning process [37].

Supervised learning can be divided into two subcategories, classification and regression. If the label of the data consist of a finite number of distinct categorical values it is a classification task. Regression, on the other hand, is used when the label does not consist of a finite number of distinct values, but rather a continuous variable [37]. Figure 2.6 displays an example of the target variable in a classification task and a regression task.

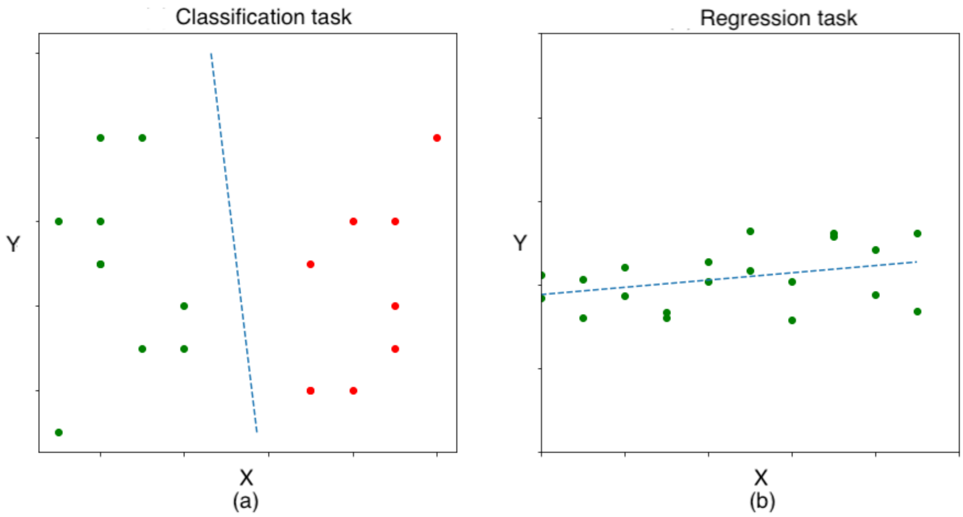


Figure 2.6: (a) displays the target variables in a classification task with a decision boundary. (b) displays the target variables in a regression task with a line of best fit.

Figure 2.5 (a) shows the learned decision boundary for a binary classification task (dotted line). The points represent the true values of the target variables for the training data. The supervised learning model learns a boundary that separates the target variables into distinct groups (line). When new data is input the predictive model, it predicts the target variables into the different groups based on the decision boundary learned during training. Figure 2.6 (b) shows the learned regression line for a regression task (line), and the true value of the target variables for the training data (points). The line represents the learned regression line, minimising the distance between the line and data points. When new data is input, the predictive model uses the intercept and slope of the learned regression line to predict the values of the target variables for the new data [37].

It is also important to note that, when choosing a supervised learning algorithm for predictions, it is important to test and compare the performance of several models. The reason for this, is the No Free-Lunch theorem by D. H. Wolpert [50]. According to the No Free-Lunch theorem, there are no a priori distinctions between supervised learners [50]. Meaning, that in general there is not one supervised learning algorithm that is in general, better than others across all problems.

2.6.2 Ordinary least squares regression:

In ordinary least squares regression (OLS) a global linear relationship between the features in the training data and a continuous target variable is modelled using a weighted linear system (Equation 2.2) [37].

$$\tilde{Y} = W^T X \tag{2.2}$$

Here $\tilde{Y}$ is the predicted target variable, X is the input data and W is the weight matrix. The structure of an OLS algorithm is displayed in figure 2.7.

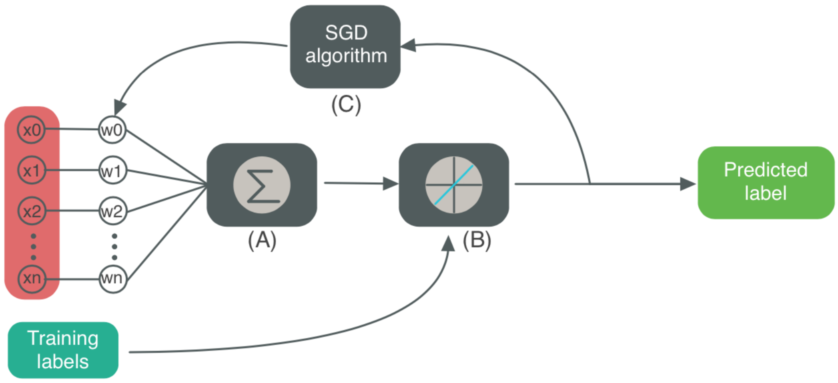


Figure 2.7: The figure displays the OLS regression model as adapted from [37]. Data is input a linear model in A. In B a cost function calculates the SSE error based on the difference of the calculated target variables from A and the true target variables in the training labels. The cost function is then minimised using an SGD algorithm in C. The model trains until the weights minimising the cost function in B are found.

Figure 2.7 shows how a sample with n features is input the OLS algorithm. In A the predicted target variable for each sample is calculated using the weighted linear system. In (B) the True target variable is compared with the predicted one to calculate the sum of squared error (SSE) for all training samples. In C, the weights, W, used in the linear system are optimised using an optimisation algorithm, such as stochastic gradient descent, to find the weights that minimise the SSE found in B. When the optimal weights are found the model is done training. The model can then be used to predict the target variable for new data using the optimised weights [37].

For nonlinear relationships between the training samples and the target variables it is possible to use polynomial regression. The learning process is the same, but polynomial terms are added to the linear system [37].

2.6.3 K-nearest regression

K-Nearest Neighbor (KNN) regression utilises the lazy learning algorithm KNN. It is known as a lazy learner because it does not learn a function to model the relationship between the input data and the target variables. Instead, during training, the algorithm memorises the input dataset [37].

When predicting the target variable for a new data sample, the KNN algorithm uses a distance metric, usually euclidean distance, to find the K most similar samples in the memorised dataset, its K nearest neighbors. The predicted value is calculated as the mean of the neighbors' target variables (Equation 2.3).

$$\tilde{y} = \frac{1}{K}\sum_{i=1}^{K} y_i \tag{2.3}$$

Here $\tilde{y}$ is the predicted target variable of a new sample, K is the number of samples that is used, and y_i is the target variable for one of the found neighbors in the memorised dataset [37].

It is crucial for the performance of the KNN algorithm to find a suitable value for K. If K is too small the algorithm is prone to overfitting and could struggle to adapt to new data, and if K is too large the model would underfit and could struggle to find a pattern in the data [37].

2.6.4 Decision tree regression

In decision tree (DT) regression, instead of modelling a global relationship between the input samples and the target variables, it models multiple piece-wise linear relationships. In DT regression the input data is split into smaller subsets of data using a DT algorithm. During training, the DT algorithm learns the optimal way of splitting a dataset with a binary decision task. The optimal split is the one that minimises the impurity (I) across the resulting subsets. The impurity of a subset is defined as the variance of the target variables associated with the samples in the subset (Equation 2.4) [37].

$$I_s = Var(y_s) = \frac{1}{M_s}\sum_{i=1}^{M}(y_i - \bar{y}_s) \tag{2.4}$$

I_s is the impurity of subset s, y_i is the true target variable of a sample in subset s and $\bar{y}_s$ is the mean of all M target variables in subset s. This splitting process is iteratively repeated, further splitting the subsets into new smaller subsets, until there is one sample left in a subset, the impurity of a subset is zero, or the impurity is less than a defined limit. An example of a trained decision tree is displayed in Figure 2.8.

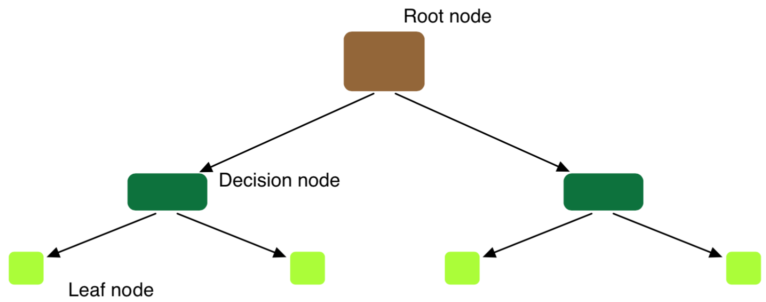


Figure 2.8: An example of a decision tree with two learned splits. Each split is done minimising the impurity across the nodes. Target variables of samples in each leaf node is used for predictions.

The first split of the input data occurs in the root node. Further splits are referred to as decision nodes. The subsets found by the final splitting is referred to as leaf nodes. The mean value of the target variables for the samples in the leaf nodes are used for predictions when the algorithm is input new data (Equation 2.5) [39].

$$\tilde{Y} = [\bar{y}_1, \bar{y}_2, ...\bar{y}_s] \tag{2.5}$$

When training a DT it is important to be wary of the number of splits. More splits makes a DT prone to overfitting [37].

2.6.5 Random forest regression

Random forest (RF) regression uses an ensemble technique to combine multiple DTs. Hence, the predicted target variables from a RF regression algorithm are calculated as the average predicted value from all the DTs (Equation 2.6) [39].

$$\tilde{Y} = [\frac{1}{P}\sum_{j=1}^{P}(\bar{y}_{1,j}), \frac{1}{P}\sum_{j=1}^{P}(\bar{y}_{2,j}), ..., \frac{1}{P}\sum_{j=1}^{P}(\bar{y}_{s,j})] \tag{2.6}$$

Here $\tilde{Y}$ is the predicted target variables for the samples input the RF model. P is the total number of DTs in the RF model, and s is the number of leaf nodes.

The RF algorithm usually has a better generalisation performance than a DT, and is less prone to overfitting [37].

2.7 Performance metrics

When evaluating the performance of supervised learning algorithms it is important to verify how well a model adapts to new data, and in general, how accurate or precise the predictions are. To evaluate how well a supervised regression model is able to model a relationship between the input data and its target variables one can use the coefficient of determination (R2-score) . R2 is given as a value between 0 and 1. A score of 1 indicates that the regression model is able to approximate the true target variables perfectly. The R2 score is calculated as the SSE divided by the Sum of total squares (SST) (equation 2.7).

$$R2 = \frac{SSE}{SST} = \frac{\sum_{i=1}^{m}(y_i - \tilde{y}_i)^2}{\sum_{i}^{m}(y_i - \bar{y}_i)^2} \tag{2.7}$$

Thus, it gives a measurement of the proportion of the SSE for predictions that is explained by the variation in the true target variable.

To investigate the overall precision of predictions the Mean Absolute Error (MAE) can be used (Equation 2.8).

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|y_i - \tilde{y}_i|. \tag{2.8}$$

MAE is calculated as the average total error between the predictions ($\tilde{y}$) and the true value target variables (y) for all samples m.

3. Data overview

.

An important aspect of this thesis was to find and create an overview of some publicly available open-source datasets that provided information relevant for modelling EV charging patterns, charging demand, and EV charging behaviour. The search for relevant datasets was done by investigating what datasets had been used in articles analysing EV charging behaviour, and by an online search using google.

The data search identified four datasets, free to use, that provided information about EV charging sessions at charging stations. Table 3.1 gives a general overview of these datasets.

Table 3.1: Table shows the different datasets investigated in this thesis. The location column gives the area where the data is recorded, the Dataset column gives the name of the dataset and the Type column gives the type of data recorded, historical charging data or real-time data.

Open source datasets

Location	Dataset	Type
California, USA	ACN dataset	Historical
Netherlands	Elaad dataset	Historical
Ireland and Northern Ireland	Ireland and Northern Ireland charging data	Historical
Norway	NOBIL database	Real-time

The datasets provide information about public charging stations in the areas shown in Figure 3.1. These four datasets are the ACN dataset from Adaptive Charging Networks (ACN) in California [22], the Elaad dataset provided by ElaadNL in the Netherlands[11], the Ireland and Northern Ireland charging data [4], and the NOBIL database from Norway [21]. All four datasets provide information about the number of EVs connecting and charging patterns at charging stations. The information, however, varies between them and the datasets have different structures. In the next sections a more detailed description of the different datasets information, and structure is presented.

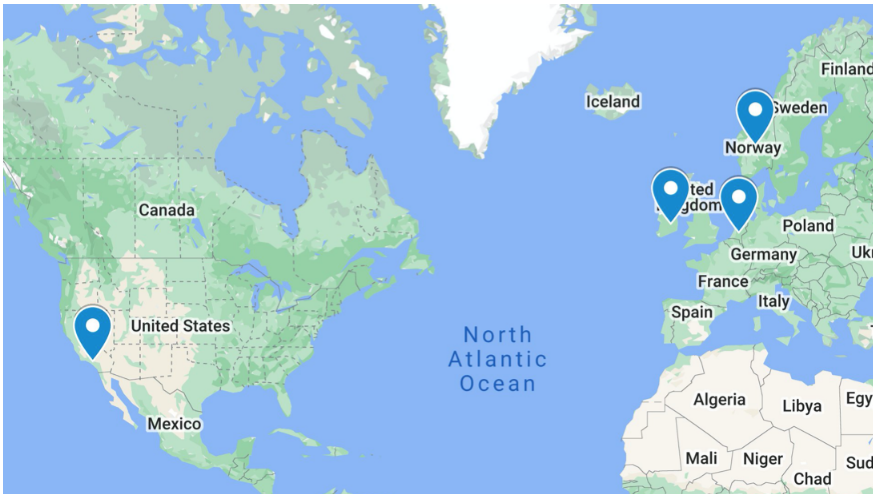


Figure 3.1: Map displaying the location of the charging stations the four datasets were recorded. The ACN dataset in California, the Elaad dataset in the Netherlands, the Ireland and Northern Ireland charging data and the NOBIL data in Norway.

3.1 The ACN dataset

The ACN dataset is available to download through a web interface and through an Application Programming Interface (API) [22]. The dataset was created and released to help researchers gain a better understanding of charging patterns and user-behaviour of public and workplace charging stations [22]. It includes historical data containing information about transactions between EVs and charging stations from two locations in California, USA. Since its inception in 2018 more than 32 000 charging sessions have been recorded and the dataset is continuously updated. The data is collected from two adaptive charging networks (ACNs) located at the California Insititute of Technology (Caltech) and at Jet propulsion laboratories (JPL) in California, USA [22].

When downloaded from the web interface the data is given as a JSON-file. The JSON-file contains two tables. One called Meta giving information of the location, start date and end date of the downloaded dataset. The second table called Items contains all the transactions between EVs and the charging station (charging sessions) at the given location. All recorded sessions are performed on type 2 charging stations.

Each sample in the dataset provides information about individual charging sessions. It includes timestamps for connection and disconnection, amount energy delivered, unique ids for each session and charging point, and user-inputs. The user-inputs are given by users when connecting to the ACN [22]. They include a unique user id, requested amount of energy, requested amount of miles, requested disconnection time, and if they paid to charge. The structure of the ACN dataset as downloaded through the web interface is shown in Table 3.2.

Table 3.2: The structure of the ACN dataset. The left columns display the features of the dataset and the right displays an example sample with placeholder values. The values in sample id, session id and user id are unique long strings containing random numbers and letters. For simplicity they are represented by -. Modified at gives the timestamp for when the user-inputs were given.

The ACN dataset

Features	Example
Sample id	-
Cluster id	0011
Connection time	Wed, 25 Apr 2018 11:08:04 GMT
Disconnect time	Wed, 25 Apr 2018 13:20:10 GMT
Done charging time	Wed, 25 Apr 2018 13:21:10 GMT
Energy delivered [kWh]	8.000
Session id	-
Site id	0002
Space id	CA-496
Station id	1-41-13-456
Timezone	America/Los Angeles
User inputs	
User id	-
Wh per mile	500.0
Energy requested [kWh]	55.5
Miles requested [$miles$]	150.0
Minutes available [min]	500.0
Modified at	2018-04-30 15:08:54
Requested departure	2018-05-01 00:17:49

Overall, the ACN dataset gives a lot of information for analysing the user-behaviour and charging pattern at a public charging station. With the transaction data it is possible to track the number of charging sessions at the ACN and the EV charging load at an aggregated station level, for periods of different time scales. Furthermore, with the addition of user-inputs the dataset makes it possible to analyse the charging behaviour of individual EV owners. The user-inputs also allows for an investigation of accuracy of user-inputs and how user-inputs affects the scheduling for a smart charging scheme. Additionally, it makes it possible to analyse how well the service provided by the ACN to EV owners is. The amount of data available also enables the dataset to be used for machine learning. For example it can be used to build models for modelling EV charging patterns, EV charging load and duration of charging sessions.

3.2 Elaad dataset

The ElaadNL data dashboard provides information to analyse electric vehicles and the way they charge in the Netherlands. According to Elaad [11], their data dashboard contains more information than any EV charging database in the world. The Elaad dataset is available through the data dashboard and includes historical transaction data from charging stations in the Netherlands. It is possible to download a sample dataset straight from the dashboard. The sample dataset contains information about 10 000 randomly selected charging sessions from 850 public charging stations in the Netherlands in 2019 [11]. It is also possible to gain access to larger dataset with more than a million charging sessions, [39]. This is done by contacting Elaad. Unfortunately, access to the larger dataset was not granted while writing this thesis. Therefore, the description done in this thesis is based on the smaller sample data. Assuming the structure of the sample data is the same as the larger dataset, the sample data provides a picture of the information available from Elaad's open data dashboard.

The sample dataset is given at a XLSX-format and it is divided into two tables. One of the tables contains historical charging session data and is called Transaction data. The other table, referred to as the Meter data, contains power and energy readings for each of the recorded sessions. In the sample dataset all recorded charging sessions were performed at public charging stations, mainly offering type 2 charging. However, some of the recorded sessions were performed with a charging level of 22 kW suggesting access to type 2 FastAC charging.

The structure of the transaction data is similar to the ACN data. It includes timestamps for connection and disconnection, energy demand and unique ids for stations, sessions and users. The dataset also includes a feature recording the average power used during the session. Table 3.3 displays the structure of the transaction data as retrieved from the open data dashboard. The Meter data table supplements the transactions data with power and energy measurements taken during each of the charging sessions. The measurements are taken at either a three or five minute interval depending on the charging station. The structure of the meter data table is given in Table 3.4.

Overall, the Elaad dataset gives an insight into the charging activity and energy demand at charging stations in the Netherlands. By combining information from the Transaction data and the Meter data, it is possible to both investigate the charging demand, number of EV connecting at public charging stations, and individual EV user-behaviour with a high resolution. The Transaction data gives the ability to track the number of EVs connecting and energy demand over different time scales, both for a charging station and at an aggregated national level. With the meter data it is possible to track the charging demand throughout a day with a high resolution, and create a load curve for a charging station. The dataset also provides information of tracking individual EV user-behaviour with the StartCard feature. Additionally, as the dataset provides information about several charging stations it is possible to compare the charging activity, charging demand, and individual EV user-behaviour across different stations. Unfortunately each station is only identified by a unique string, hence the location of a station is unknown.

Table 3.3: The structure of the Elaad transaction data. The left columns display the features of the dataset and the right displays an example sample with placeholder values. The values in Transaction id, Charge point and StartCard are unique long strings containing random numbers and letters. For simplicity they are represented by -.

The ELAAD transaction data

Features	*Example*
Transaction id	-
Charge point	-
Connector	1
UTC Transaction Start	2019-01-01 11:49:04
UTC Transaction Stop	2019-01-01 13:10:10
StartCard	-
Connected time [*hour*]	2.39
Charge time [*hour*]	2.10
Total energy [kWh]	3.34
Max power [kW]	3.242

Table 3.4: The structure of the Elaad power meter data. The left columns display the features of the dataset and the right displays an example sample with placeholder values. The values in Transaction id and charge point are unique long strings containing random numbers and letters. For simplicity they are represented by -.

The ELAAD Power Meter data

Features	*Example*
Transaction id	-
Chargepoint	-
connector	1
UTCTime	2019-01-01 11:49:04
Collected value	5394520
Energy interval [kWh]	0.89
Average power [kW]	3.242

3.3 Ireland and Northern Ireland dataset

The Ireland and Northern Ireland charging data was made by J. Burkin, and is available to download through a web interface [4]. The dataset provides status updates for charging points, from around 1100 charging stations in Ireland and Northern Ireland, given at five minute intervals. It is based on real-time data provided by ESB E-Cars charge map [12]. The dataset contains charging data from the period November 2016 to June 2019 [4].

The Ireland and Northern Ireland dataset is available to download as multiple txt-files. Each txt-file contains all the recorded measurements from one month. All the separate txt-files have the same structure. There are three types of charging stations included in the dataset, type 2 one-phase, type 2 FastAC, and Fast D.C. charging stations.

Each sample, in the dataset gives a timestamp of a sensor status update, the location of the charging station (both address and coordinates), type of socket, and the sensor status. There are four possible sensor statuses included in the dataset. These are Fully occupied (Occ), partially occupied (par), out of service (Oos), and out of contact (Ooc). When a station is vacant, there is no recorded measurement, thus when a station is vacant can be identified by the lack of a measurement [4]. The structure of the dataset is displayed in Table 3.5.

Table 3.5: The structure of the Ireland and Northern Ireland dataset. The left column gives the features in the dataset and the right column gives an example sample with placeholder values. Charge point id is a unique string for each charge point. It is represented by - for simplicity.

The Ireland and Northern Ireland charging data

Features	*Example*
Date	20190701
Time	1
Charge point id	-
Charge point type	StandardType2
Status	Occ
Values	-6.923077,52.841045
Address	The Parade, Bagenal...
Longitude	-6.923077
Latitude	52.841045

Overall, the Ireland and Northern Ireland dataset by J. Burkin [4], gives an insight into EV charging at public charging stations in Ireland and Northern Ireland. Using the charge point status it is possible to track the the number of EVs connecting to the charging stations. With the connector type and charge level it also provides some insight into energy demand at the charging stations. Furthermore, as each measurement is assigned to distinct stations with given locations it is possible to analyse differences between charging stations. Additionally, the dataset contains a vast amount of samples making it eligible to use for machine learning. E.g. the dataset was made for the purpose of predicting charge point availability [4].

3.4 NOBIL database

The NOBIL database is an open, central, and publicly owned database. It was created to secure an overview of the Norwegian charging infrastructure, provide valuable information to EV owners, and developers who aim to create useful tools for EV owners [21]. The database provides information about more than 2500 charging stations in Norway. Unlike the ACN and Elaad datasets, the NOBIL database does not store historical transaction data, but offers a description of the charging infrastructure at stations, and real-time information from some charging stations. Real-time information is stored in the database for up to 7 days. Thus, with permission from NOBIL, it could be possible to create a database by regularly downloading real-time information, similar to the Ireland and Northern Ireland dataset by J. Burkin [4].

Data is available to download from the NOBIL database through an API. The API is free to use, but one must register as a user at NOBILs web page to receive an API-key. The API allows for quite advanced requests to retrieve desired information. Specific methods that are available through the API are described in the API user-manual, available at the NOBIL web page [34]. An overview of the attributes available in the NOBIL database are available online at [32]. With a specified request, it is possible to download data that contains real-time information from several charging stations and information of their infrastructure through the API.

Table 3.6 shows an example of data that can be downloaded through the NOBIL API. With a specified request it is possible to to retrieve information about the location, type of charging available, and real-time information. The real-time information is similar to that of the Ireland data, in that it provides a timestamp, and connector status for each of charging points at a station. The available statuses are Vacant and Busy.

Overall, with data displayed in Table 3.6, it is possible to analyse charging patterns at public charging stations in Norway. With the connector statuses and connector indexes it is possible to track the number of charging sessions at a charging station. With the connector type and charge level it is also possible to gain some insight into the charging demand from the charging stations. Furthermore, using the connector index it should be possible to separate between different charging sessions, and analyse charging session duration. Furthermore, with the connector status and charge mode it could be possible to analyse how often EV owners leave their EV connected after it has finished charging. The known location of each charging station also makes it possible to compare the activity between areas, e.g. inner-city and rural areas.

Table 3.6: The structure of data from the NOBIL database as retrieved through the API. The left column shows the features retrieved through the API query and the left and example sample with placeholder values.

NOBIL data

Features	*Example*
Province	Trøndelag
City	Trondheim
Street	Haakon Vii gate
Street #	17
Connectors	8
Connector index	3
Connector type	CHAdeMO
Charge level	50 kW - 500VDC max 100A
Timestamp	2021-03-10T06:45:07
Connector status	Vacant
Charge mode	Mode 4

3.5 Other datasets

A more comprehensive overview of open source datasets is included in the study 'A review of Electric Vehicle Load Open Data and Models' conducted by Y. Amara-Ouali et al. [1]. The study investigated more than 860 data repositories containing charging station data. The study found 60 data repositories with information relevant for modelling EV charging load from 15 different countries. Like, this thesis the study focused on transaction data between EVs and EVSEs. However, it also included datasets with other variables, such as traffic, travels surveys, and air quality [1].

4. Method

The analysis was conducted using the ACN dataset [22]. The historical transaction data in the ACN dataset allowed for an analysis of charging patterns, and aspects of EV user-behaviour at a smart charging station. The main aspects of EV user-behaviour that was investigated in this analysis was, the time of connection, charging session duration, and energy demand from charging sessions. Moreover, as the ACN data included user-inputs, which were used for smart scheduling, it was analysed how well the ACN was able to meet the users demand, and overall if the user-inputs were accurate. In addition to the data-driven analysis of user-behaviour, it was attempted to predict aspects of user-behaviour, such as charging session duration and energy delivered in a charging session, using supervised learning regression.

Data spanning the period 25. April 2018 until 1. January 2021 was downloaded through the web interface. The downloaded dataset contained 32 307 samples. The dataset had two main types of samples. Historical transaction data with user-inputs (claimed sessions) and historical transaction data without user-inputs (unclaimed sessions). It was decided to only use data from the Caltech ACN because the station was representative of both a workplace charging station and a public charging station [22].

The data processing, analysis and machine learning was done using Python [46]. The downloaded data was unpacked into a format called Pandas DataFrame, a tool for data management in python [44], [26]. Visualisation was done using tools from pandas [44], [26], matplotlib.pyplot [17], and Seaborn [48]. For the machine learning, tools and models available through the Sci-Kit learn library were used [36].

4.1 Data processing

Before the dataset was used for the analysis, some pre-processing was done. The pre-processing included detection of duplicated data, detection of missing data, detection of outliers, and feature engineering.

4.1.1 Duplicated data

To identify if there was duplicated data in the downloaded dataset, the unique session ids provided in the dataset was used [22]. If a session id was repeated for several samples in the data, it suggested that there was duplicated data in the dataset. Overall, there were 32 307 samples in the downloaded dataset, but only 28 391 unique session ids. Further investigation found that the samples with the same session id provided the same information in the transaction data, but some values in the user-inputs varied for some of the claimed sessions. To prevent the results from being twisted or weakened by duplicated samples, it was decided to remove them. For unclaimed sessions the first appearance of a session id was kept, and for the claimed sessions the sample with the most accurate Energy requested compared to the Energy delivered was kept. If the input Energy requested was equal for all occurrences of a session id the first occurrence was kept.

4.1.2 Detection of missing data

Missing values in a dataset are represented by a Non-value (NaN) in Python. According to S. Raschka and V. Mirjalili [37], it is important to identify and replace missing values since most computational tools are unable to handle them, and they can produce errors and unpredictable results.

The downloaded dataset included missing values in the user-inputs for unclaimed sessions, as expected. However, in addition to missing data in the user-inputs of unclaimed sessions, there where 1152 other values missing from the dataset. These were all in the Done charging time feature. The Done charging time feature records the timestamp of when the last non-zero current was drawn from the EVSE [22]. A missing value in this feature suggested that the vehicle had disconnected before the EV was fully charged, or before the Energy requested was delivered.

It was decided to impute values from the Disconnect time feature to replace the NaN-values in the Done charging time feature. The reason for this, was that a missing value in the Done charging time feature indicated that the EV had disconnected, before it was done charging. Hence, charging the entire period until it disconnected. On the other hand, the missing values in the user-input features for unclaimed sessions were not changed as these data points were not included in any machine learning algorithm, nor were unclaimed sessions used for analysing user-inputs.

4.1.3 Outlier detection

An outlier is a value in a feature that deviates from the feature mean by an abnormally large amount. Hence, it is important to investigate data for outlier values as they can distort the distribution of the features, and prevent trends and patterns from being found in the data when an analysis is conducted. Some of the features in the ACN dataset were prone to contain outlier values. These features were the Energy delivered, Minutes available, Miles requested and Energy requested features. The Energy delivered feature contains the Energy delivered to an EV during a charging session. An Outlier values can occur in this feature if a faulty measurement is stored in the data. The Miles requested, Minutes available, and Energy requested features can contain outliers if an EV owner has given inaccurate values when providing the user-inputs to schedule their charging session. The other features in the dataset either contain boolean values, unique strings or timestamps, thus, they should not include outliers.

Figure 4.1 displays the distribution of the Minutes available, Energy delivered, Energy requested, and Miles requested features.

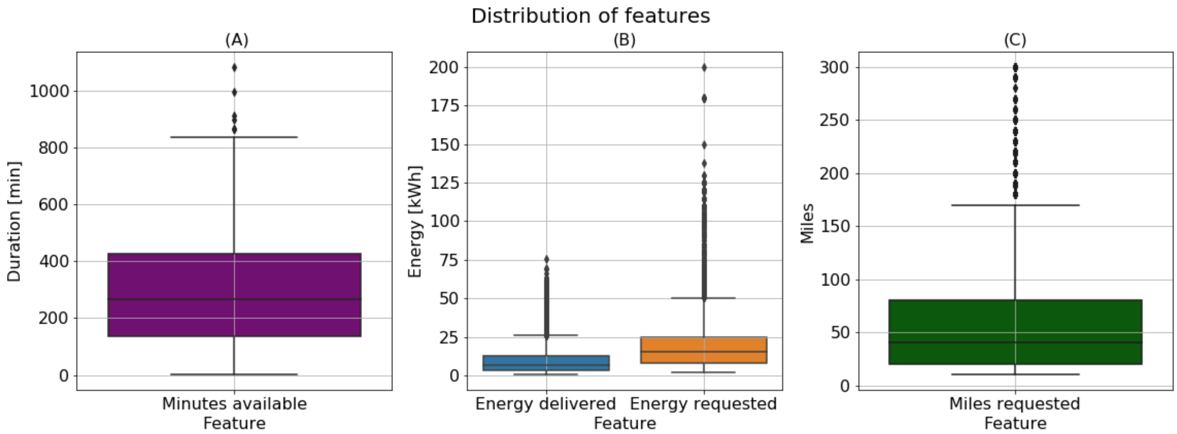


Figure 4.1: Distribution and characteristics of Minutes available (A), Energy delivered and Energy requested (B), and Miles Requested (C).

Figure 4.1 (A) shows the distribution of the Minutes available feature. From the figure it was found that most recorded values were between 150 minutes and 430 minutes. (A) also identifies that there were large outliers, where EV owners had stated a long period for the EV to be connected. Figure 4.1 (B) display the distribution of the Energy delivered and Energy requested features. Most recorded values were between 5 kWh and 10 kWh, and 10 kWh and 25 kWh for Energy delivered and Energy requested, respectively. The figure identifies outliers in both features. Moreover, it displays very large outliers recorded in the Energy requested feature. Figure 4.1 (C) displays the distribution of values in the Miles requested feature. Most of the recorded values were between 10 miles and 80 miles. Additionally, there were large outliers found in the feature.

When handling outliers it is of high importance to investigate why they appear and if they can appear again. The largest value that occurred in the Energy delivered feature was 75.53 kWh. This value appears as an outlier in the data. However, there are EVs on the market with a larger battery capacities than this. Furthermore, outliers in the Minutes available, Miles requested, and Energy requested features reflect the unpredictability of EV owners. Erroneous values can appear in user-inputs if, for example, an EV owner does not know the technical aspects of charging their EV, like how many kWh is enough to charge the EV.

It was decided to not remove samples from the dataset based on outlier values in any of the investigated features. The reason for this, was that, outlier values could be expected to occur again in each of the respective features. Moreover, they provide information which could be useful in the analysis. E.g. when investigating user-inputs, an important aspect was to investigate if they were reliable. Inaccurate user-inputs giving outlier values shows that user-inputs can be erroneous which would also be an important part of the result.

4.1.4 Feature engineering

Some new features were extracted and added to the dataset. To investigate how long an EV was connected to the charging station, a Session duration feature was added. This Session duration feature presented the period an EV was connected in terms of minutes. It was calculated according to Equation 4.1

$$Session\ Duration(i) = Disconnect\ time(i) - Connection\ time(i). \tag{4.1}$$

Here, i represents a sample in the dataset. The calculation was done for all samples. Similarly a Charging duration feature was calculated using the values from the Done charging time feature instead of the connection time feature. The reason for this was to investigate how often EV owners left their car connected after it had finished charging.

The connection time feature was also modified. Instead of the connection time being represented as a timestamp, it was split into three features, giving the date, the weekday as a value between 0 and 6, and the time of connection in minutes since midnight. The main reason for this was to use the Connection time and Weekday features for Machine Learning, and investigate the average changes in charging pattern during a week. Prior to using the dataset for machine learning, the Weekday feature was further split into seven dummy features. Thus each Weekday feature gave either one or zero, depending on if the charging session was conducted on the respective day.

Investigating the engineered session duration and charging duration features identified, that both features included outliers. These features were made on the assumption that the Connection time and Disconnect time timestamps were correct. However, the identified outliers suggested otherwise.

It was decided to remove samples with extremely large values in the Session duration and Charging duration features. All samples with values deviating more than 3 standard deviations (STD) from the mean feature value were removed. This was 383 samples in total.

4.1.5 The processed dataset

After pre-processing there were 28 008 sessions left in the data, of which 13 453 were claimed sessions. Overall, 3 916 samples where removed as duplicates and a further 383 where removed due to outlier values in the engineered features. Table 4.1 shows the structure of the ACN dataset after processing.

Table 4.1: Table displays the structure of the ACN dataset after prepossessing. The left column gives the feature left in the dataset and the right column an example with placeholder values.

Processed ACN dataset

Features	*Example*
Session id	-
Connection date	2019-04-25
Weekday	2
Connection time	11:43:11
Session duration [min]	Wed, 636
Energy delivered [kWh]	27.6
User id	-
Minutes available [kWh]	648
Requested departure	2018-04-26 00:17:49
Miles requested [miles]	70
Energy requested [kWh]	28.0
Payment option	True

4.2 Identifying a charging pattern and significant events

The first part of the analysis focused on analysing the evolution of the number of connecting EVs at the ACN, identify a charging pattern for the ACN and investigate aspects of user-behaviour on a general level, using an average of all recorded charging sessions.

When analysing the evolution of connecting EVs the entire downloaded period, April 2018 to January 2021, was used. The goal was to investigate if the number of EVs at the ACN increased throughout the period, identify significant events that overall changed the charging activity at the ACN, and to investigate if there was any seasonal dependency for the connecting EVs. E.g. a varying amount of energy charged.

The analysis of the charging pattern at the ACN was conducted for a longer period, where charging activity was stable. A stable period was defined as a longer period where the number of connections was similar, and not affected by any external events seen in the data. The general charging pattern was investigated as the average number of EVs connecting in each hour.

The analysis of general patterns for the observed EV user-behaviour was conducted investigating the average Session duration and the average Energy delivered for all charging sessions, for each day of the week. This was done to investigate if there were any general trends or patterns displayed by the EV owners charging at the ACN. Moreover, if in general EV owners displayed an inherently different user-behaviour on certain days.

Lastly, as the ACN was a smart charging station. It was investigated how well the ACN was able to deliver the Energy requested by EV owners within the requested period. This was done by comparing the user-inputs and service delivered. The user-inputs were also investigated to observe how they varied from the service delivered by the ACN. If EV owners disconnected when they stated in the user-inputs, and if EV owners requested more energy than what they received.

4.3 Analysing user-behaviour in charging groups

After investigating the general charging pattern and a general user-behaviour for EV owners charging at the ACN, it was analysed if the EV owners could be separated into charging groups. Using the unique User id provided in claimed sessions, it was possible to see how many times each EV owner charged at the ACN in the period. EV owners were categorised into distinct charging groups based on how frequently they charged.

The charging groups were further used to analyse if the EV owners in each charging group displayed a distinctly different user-behaviour. This was done by comparing distributions of Connection time, Session duration, Energy delivered, and inaccuracies in user-inputs for the EV owners, in each charging group. The user-behaviour of each EV owner was represented by the average user-behaviour they displayed across all their charging sessions.

4.4 Predicting user-behaviour with machine learning

Another part of the analysis was to attempt to predict aspects of EV user-behaviour using Machine Learning. Supervised learning regression algorithms were trained and used to predict Session duration and Energy delivered based on the data available in the ACN dataset.

Five supervised learning regression algorithms were trained and tested for predicting Session duration. The models were trained on two different subsets of the ACN dataset. One subset only including information from Connection time and Weekday, and one subset including some additional information about the EV connecting. Namely, the User id and Energy requested features. Likewise, for predicting Energy delivered. Five supervised learning regression algorithms were trained and tested, using the two subsets of the ACN dataset. However, for the larger subset, instead of including Energy requested, the Minutes available feature was included.

The goal of training the machine learning models on different subsets, providing a different degree of descriptive data, was to investigate how important the time of connection was for the predictions, and how important information about the individual EV owners was.

All the machine learning models were trained and tuned using a grid search. A grid search takes a range of values for each tuneable parameter for a machine learning model, trains it, and identifies which set of the given parameters provide the best performing model [36]. Table 4.2 gives a overview of the five supervised learning regression algorithms, and their tuneable parameters, that were used for predictions in this thesis. Only a general description of the parameters are given in Table 4.2. A more in depth description is given in the Scikit-learn library [36].

To analyse and validate the performance of the tuned models when input new data, a KFold cross-validation with 5 folds was used [36]. Furthermore, the metric used to validate the performance of a model was the R2-score.

The predictions from the best performing machine learning model was then compared with the user-inputs. This was done by comparing the MAE of the user-inputs, and the MAE of the predictions. Finally, the best performing models were also investigated to identify which features had the greatest importance for the performance of the models.

Table 4.2: Displays the supervised machine learning regression models used to predict user behaviour. For each of the models, the table also gives the parameters that are tuned with a grid search, to find the best performing models.

Model:	Hyper parameters				
OLS regression					
OLS regression - polynomial	**Degree:** Order of polynomial to fit the input data.				
KNearest regression	**# neighbours:** Number of neighbors to use for kneighbors queries.	**Weights:** Weight function used in prediction.	**Algorithm:** Algorithm used to compute the nearest neighbours	**P:** Power parameter for Minkowski metric.	
Decision tree regression	**Max depth:** The maximum depth of the tree.	**Min sample split:** The minimum number of samples required to split an internal node.	**Min sample leaf:** The minimum number of samples required to be at a leaf node.	**Criterion:** The function to measure the quality of a split.	
Random forest regression	**Max depth:** The maximum depth of the trees.	**Min sample split:** The minimum number of samples required to split an internal node.	**Min sample leaf:** The minimum number of samples required to be at a leaf node.	**Criterion:** The function to measure the quality of a split.	**# estimators:** The number of decision trees in the forest.

5. Results and discussion

5.1 Identifying a charging pattern and significant events

5.1.1 Activity at ACN April 2018 to December 2020

Figure 5.1 shows the general evolution of activity at the Caltech ACN for the entire downloaded period, 25. April 2018 to 1. January 2021.

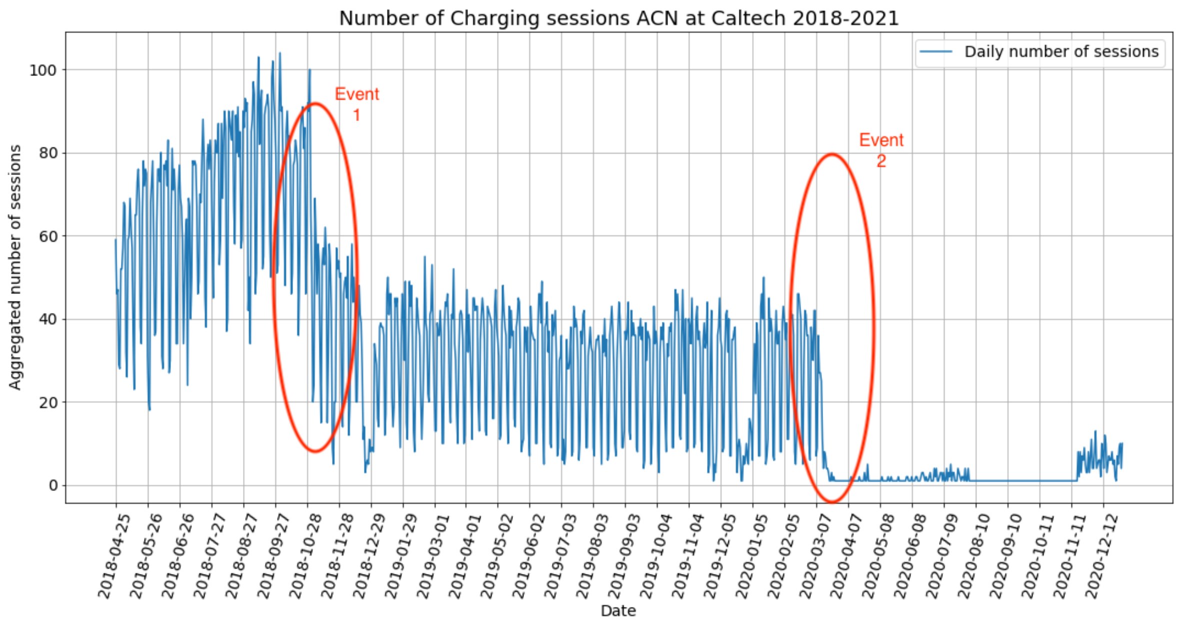


Figure 5.1: The figure shows the number of EVs connecting to the ACN per day for the entire downloaded period, 25. April 2018 to 1. January 2021. Additionally, the Figure displays two events with a long-term impact, that changed the overall number of EVs connecting to the ACN.

In the figure, the curve shows the daily number of connections at the ACN. The variation between local peaks and troughs display the difference in the number of EVs connecting on weekdays and weekends, respectively. The weekday-weekend variation is repetitive and systematic, but quite substantial. The figure also distinguishes two events that abruptly changed the charging activity at the ACN, and left a significant long-term impact (Event 1 and Event 2).

In the period April 2018 to September 2018, the number of connections per day increased, both for weekdays and weekends, as seen in Figure 5.1. For weekdays, the number of connections grew from around 60 connections per day to more than 90, peaking in August at more than 100 connections in a day. Similarly, for weekends the number of connections grew from around 30 to around 50 in August and September. Such an increase in connection was not seen at the ACN for the remainder of the downloaded period. There could be multiple reasons for why there was such a large increase in connections in 2018. One reason could be due to the electrification of the transport sector in California. Like Norway, California has seen a broad adoption of EVs over the last couple of years. Between 2013 and 2018, the number of registered EVs grew from 25 146 to 232 236. In 2018 alone, there were 72 534 new EVs registered. By the end of 2020, EVs accounted for 1.2 % of the car park in California [6]. Another reason that could explain the rapid increase of EVs charging at the ACN, was explained by Z. J. Lee et al. in 'ACN-Data: Analysis and applications of an open EV charging dataset' [22]. According to Z. J. Lee et al. [22], it was free to charge at the ACN prior to November 1. 2018. This could have incentivised EV owners to charge for free at the ACN instead of charging elsewhere.

After the period with increasing activity at the ACN, the number of connections per day at the ACN was reduced. This was marked as Event 1 in Figure 5.1. The nature of Event 1 was not explicitly described in the dataset, other than that after it happened there was a significantly lower amount of EVs connecting to the ACN. Event 1 was, however, further explained in the article 'ACN-Data: Analysis and applications of an open EV charging dataset' by Z. J. Lee et al. [22]. 1. November 2018 a $0.12\frac{\$}{kWh}$ charging fee was introduced [22]. Thus, Event 1 could show how the introduction of paid charging can affect how EV owners choose to charge.

After the introduction of the charging fee in November 2018, the number of connections at the ACN was similar until March of 2020, ignoring the weekday-weekend variation. In march 2020, Event 2 occurred. Event 2 reduced the number of EVs connecting to the ACN to almost none for the rest of 2020. Like for Event 1, Event 2 was not explicitly described in the data other than the significant reduction in connections. However, Event 2 coincided with when the COVID-19 pandemic caused a statewide lockdown in California. 19. March 2020, The Governor of California, Gavin Newsom, issued an executive order, calling for California's residents to 'Stay at home' to disrupt the spread of the COVID-19 virus [30]. This led to Caltech closing its doors for everyone except essential workers on March 19. 2020 [38]. Thus, it could be assumed that Event 2 and the following period shows how the ACN was affected by the lockdown caused by the pandemic.

In addition to Event 1 and Event 2, two short-term periods with a reduced number of EVs connecting were observed in Figure 5.1. These periods were in December 2018 and December 2019. During these periods, the overall number of EVs connecting to the ACN on weekdays dropped to a similar level to that of weekends, but only for a short time and with no significant long-term impact. The reason for these events were not explained in the data, but with the relevant dates, they were identified to be during the Christmas holidays, with the Caltech Academic Calendar [7]. Thus, it seems reasonable to assume that the reduction in EVs charging at the ACN was a consequence of the Christmas holiday.

It was also investigated how the energy demand at Caltech varied throughout a year. Using data from 2019, it was possible to investigate how the aggregated energy demand varied for each month. Figure 5.2 shows the monthly aggregated energy demand at the ACN in 2019.

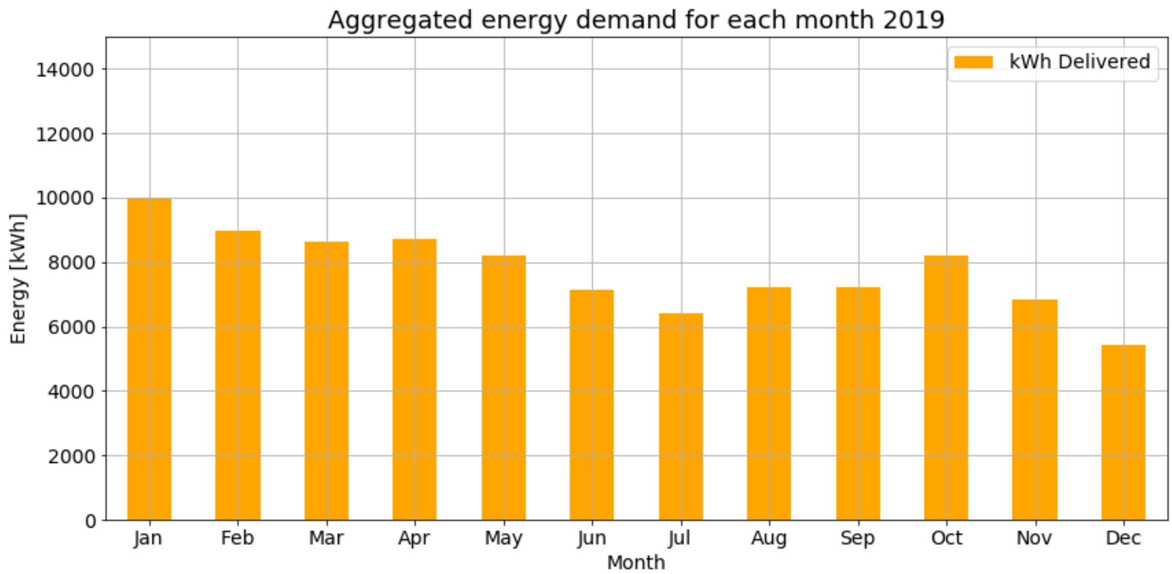


Figure 5.2: Aggregated monthly energy demand for the ACN in 2019.

Figure 5.2 shows that there was some variation in energy demand during the different months of the year. The highest energy demand was recorded in January, with an aggregated energy demand of around 10 000 kWh. The aggregated energy demand dropped towards the summer months, and in July there was an aggregated energy demand of around 6 500 kWh. During the fall months, the aggregated energy demand increased, but a reduction was seen in November and December. An explanation for the variation in energy demand could be the variation in daily number of connections as seen in Figure 5.1. However, a larger energy demand in winter months, like January and February, than for a summer month, like July, could also be a consequence of how a colder ambient temperatures affected the driving range and battery consumption of the

EVs. Studies investigating the effects of ambient temperature on distinct EV models have found that a colder ambient temperature could increase the power consumption per kilometre for some EV models. E. g. a study by X. Hao et al. [15], and a study by T. Yuksel and J. J. Michalek [51]. Hence, assuming the EV owners used their EV similarly for all months of 2019. A colder ambient temperature in the winter months could have forced the EV owners to charge their EV with more energy to be able to reach their desired destinations.

5.1.2 Electric vehicles charging in 2019

The further analysis investigating a daily charging pattern and the general EV user-behaviour was conducted using only charging sessions from 2019. The reason for this, was to allow for an investigation of how, on average, the EV owners charged at the ACN, and prevent the impact of Event 1 and Event 2 from affecting the analysis. Figure 5.3 shows the daily number of connections and average daily connections for each week in 2019.

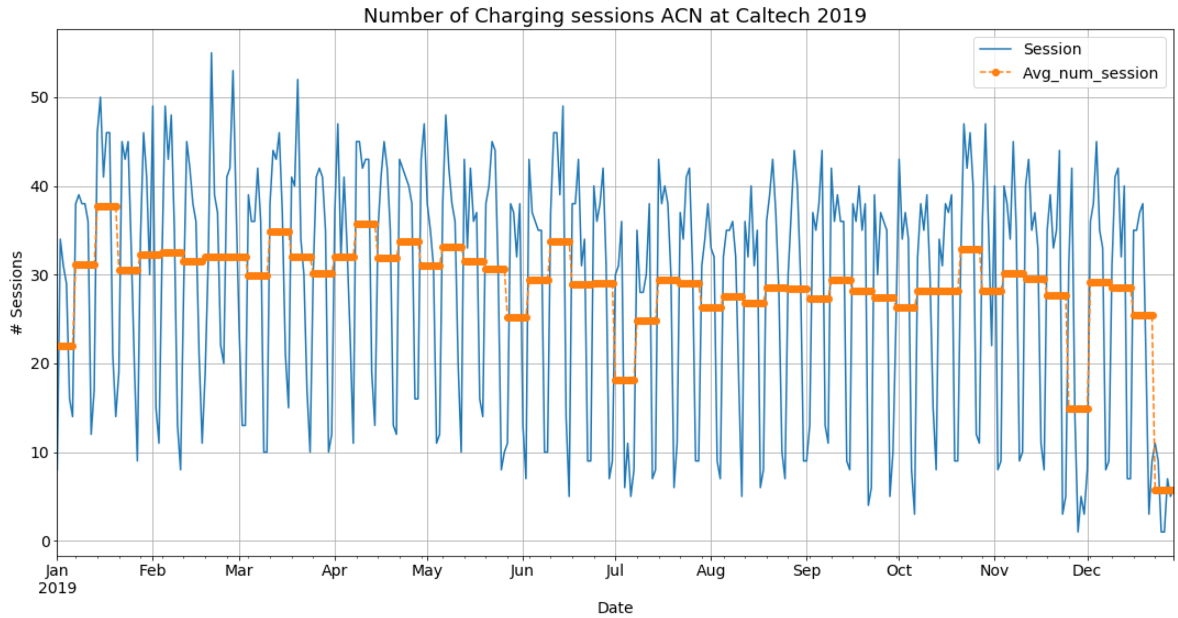


Figure 5.3: The daily number of EVs connecting and weekly mean number of connections per day at the ACN in 2019.

As seen in Figure 5.3, the weekly charging pattern was similar for most of 2019. There were a large amount of EVs charging on weekdays, and significantly less EVs charging on weekends. On average, around 40 EVs charging every weekday, and around 10 EVs charging on Saturdays and Sundays. There were, a couple of periods where the number of connections on weekdays was significantly reduced, identified by the weekly average in Figure 5.3. Namely, the first week of January, first week of July, and the first and last week of December. With the Caltech Academic calendar, these periods were identified as holidays [7].

5.1.3 Daily charging pattern

Figure 5.4 shows the average daily charging pattern for the ACN in 2019.

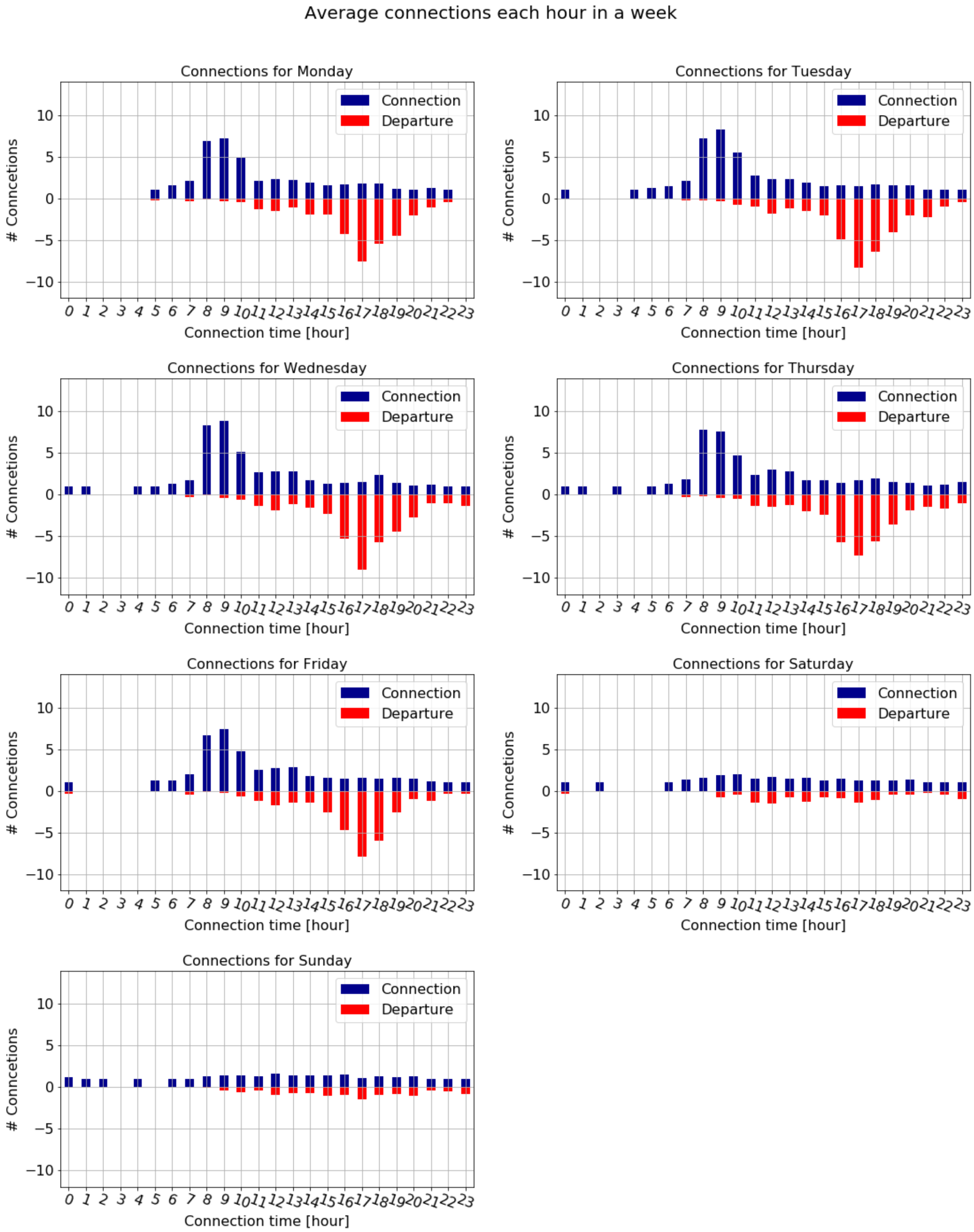
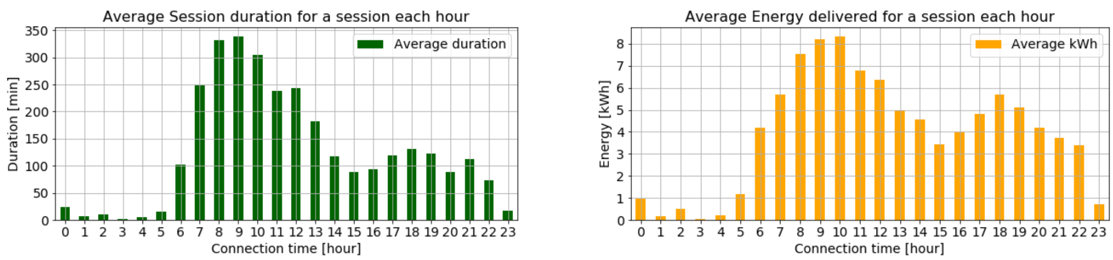


Figure 5.4: Daily charging pattern at the ACN, based on the average number of connections and disconnections each hour.

Figure 5.4 shows the average number of connections and disconnections each hour for all the days of the week. In the figure, each bar represents the average number of connections and disconnections between the given timestamp and the next, E.g. 8 gives the average number of connections and disconnections between 8 and 9. Connections are represented by a positive value, as it can be seen as EVs being added to the charging station, while disconnections are represented by a negative value, as they result in EVs being removed from the charging station.

According to Z. J. Lee et al. [22], the ACN at Caltech was operated as a combined workplace and public charging station. This was also seen when investigating the average daily charging pattern, as seen in Figure 5.4. It was found that there were significantly more EVs connecting on weekdays than on weekends. Moreover, it was found that the distribution of the connections throughout the day, the charging pattern was distinctly different. On weekdays, there was a peak of connecting EVs in the morning between 8 and 10. On average, there were 6 to 9 EVs connecting in each of these hours, each weekday. The peak in connections were followed by a peak in disconnections between 17 and 18. This pattern was similar to the expected charging pattern for a workplace charging station presented in Figure 2.2. In addition to this, there were some EVs connecting during the other hours of the day, with a slight increase in connections around 17 to 18. This suggests that some EV owners used the ACN as a public charging station. As shown in Figure 2.2, the expected charging pattern for a public charging station has a small peak in the morning similar to that of a workplace charging station, but also a small peak in connections in the evening. On weekends on the other hand, there was no distinct peak in the morning, and the connecting EVs arrived more evenly distributed throughout the day. From Figure 5.4, it can be seen that around 1-2 EVs connected each hour during the day, evening and night. The charging pattern observed on weekends suggests that the ACN was used more as a public charging station on weekends.

Figure 5.5 displays the average Session duration and Energy delivered for charging sessions conducted in different hours of the day on a weekday.



(a) Average Session duration for a charging session performed each hour of a day.

(b) Average Energy delivered for a charging session performed each hour of a day.

Figure 5.5: Average Session duration and Average energy delivered for a charging session conducted during different hours.

Figure 5.5 (a) shows the average session duration of charging sessions in 2019 based on their connection time. Figure 5.5 (b) display the energy delivered for charging sessions based on when the EV connected to charge.

In Figure 5.5 (a), it is observed a general trend that EVs connecting between 8 and 10 are left connected for longer than EVs connecting later in the day, and the EVs are often connected for a duration of about 320 minutes to 350 minutes. This could suggest that many of the EV owners leave their EV connected during their workday and only disconnect when they leave work. In Figure 5.5 (b), it is observed that EVs connecting in the morning between 8 and 10 usually received more energy than EVs connecting in the other hours of the day. This could be a consequence of EV owners who charge in the morning leave their EV to be fully charged, while EVs connecting later in the day only charge the amount necessary to reach another destination.

5.1.4 Average charging session duration

Figure 5.6 shows the average Session duration, and the average Charging duration for charging sessions in 2019.

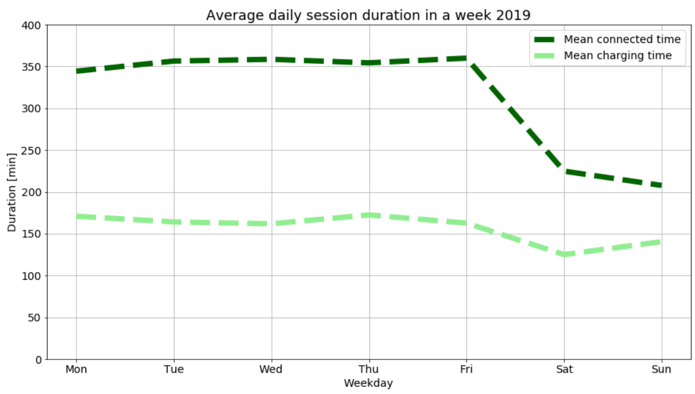


Figure 5.6: The average charging session duration and average charging duration for charging sessions performed at the ACN.

The figure shows that on weekdays, the average session duration was around 350 minutes. Of this connected time, only about half was spent charging. On weekends, the average charging session duration was significantly lower. The charging duration also decreased on weekends, but only slightly.

Table 5.1 further shows the average charging session duration, and the STD for charging session duration for each day of the week.

Table 5.1: Average charging session duration and variation in charging session duration, for each day of the week.

Variation in charging session duration

	Mon.	*Tue.*	*Wed.*	*Thu.*	*Fri.*	*Sat.*	*Sun.*
Mean [min]	344	357	359	354	360	225	208
STD [min]	43.5	45.6	36.6	41.8	58.4	80.9	69.1

Table 5.1 shows that there was quite a large variation for charging session duration for the EVs charging at the ACN. Moreover, a greater variation was seen on weekends than on weekdays. The smaller variation on weekdays, however, could be a consequence of the majority of EV owners connecting on weekdays used the ACN as a workplace charging station. Hence, the charging session duration was more dependent of the work hours of the EV owners, rather than the time needed to charge the EV. The variation was still quite substantial, which could be due to EV owners working different hours or that the charging station also was available as a public charging station. As seen on weekends, when the ACNs charging pattern was mostly representative of a public charging station, there was a large variation in the charging session duration. This could imply that charging sessions conducted at a public charging station are more specific to the energy needs of the EV owner.

5.1.5 Average energy delivered per session

Figure 5.7 shows the average energy delivered in a charging session in 2019.

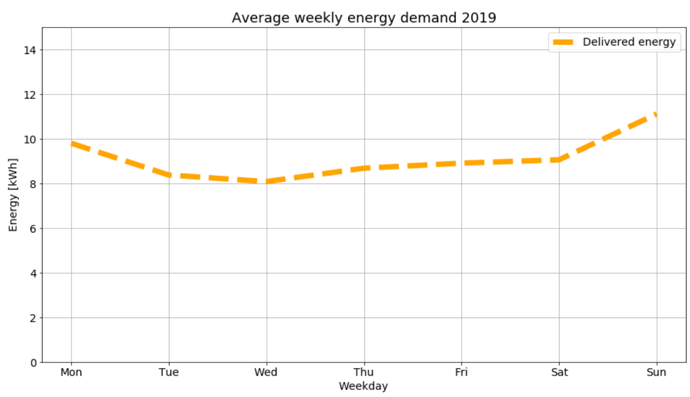


Figure 5.7: Average energy delivered for charging sessions performed at the ACN.

As seen in Figure 5.7, the energy delivered in a charging session on average ranged between 8 kWh and 11 kWh throughout the week, with the highest averages found for Sundays and Mondays. Table 5.2 further display the average energy delivered in a charging session and the STD for each day of the week.

Table 5.2: Average Energy delivered and variation in Energy delivered per charging session, for each day of the week.

Variation in Energy delivered

	Mon.	*Tue.*	*Wed.*	*Thu.*	*Fri.*	*Sat.*	*Sun.*
Mean [kWh]	9.8	8.4	8.1	8.7	8.9	9.1	11
STD [kWh]	1.6	1.6	1.8	2.5	1.7	3.2	3.8

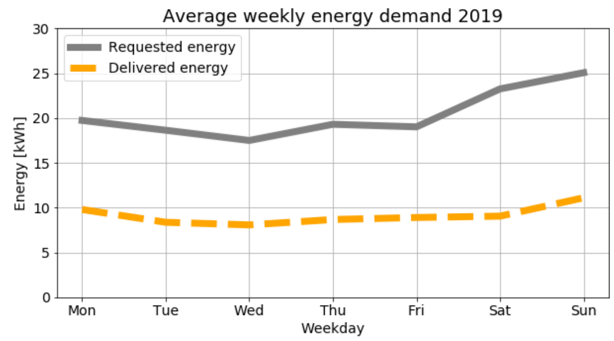
As shown in both Figure 5.7 and Table 5.2, the charging sessions where the most energy was delivered were performed during Mondays and Sundays. The lowest average energy delivered were recorded on Tuesdays and Wednesdays. Table 5.2 further display that there was a larger variation in Energy delivered for charging sessions performed on weekends, than on weekdays. On weekends, there was a STD of 3.2 kWh and 3.8 kWh, on Saturdays and Sundays respectively. This variation it could indicate that EV owners charging on weekends display a more diverse charging behaviour than the EV owners charging on weekdays. This could reflect the different charging patterns seen for weekdays and weekends at the ACN. The observations in Figure 5.6 suggested that on weekdays a majority of the connecting EV owners left their EV to charge when going to work, while on weekends a majority of EV owners connect to the ACN solely for the purpose of charging. As discussed when investigating the charging session duration, the large variation on weekends could be a consequence of the access EV owners had to charge elsewhere. EV owners charging a large amount of energy on weekends could have a limited selection of locations to charge, hence have to charge their EV fully at the ACN. EV owners charging a low amount of energy might only need to charge their EV to reach another destination, charging their EV fully elsewhere. In addition, fewer charging sessions were recorded on weekends, hence the diversity of energy demand became more prevalent on weekends, as indicated with the larger STD in Table 5.2.

To get a deeper understanding of why there is such a large variance in Session duration and Energy delivered seen in the data, a more thorough understanding of each EV owner is needed. E.g. where and how they live, what type of EV they drive, and in general a better understanding of their driving habits. Unfortunately this is not recorded in the ACN dataset beyond the distinct User id in claimed sessions.

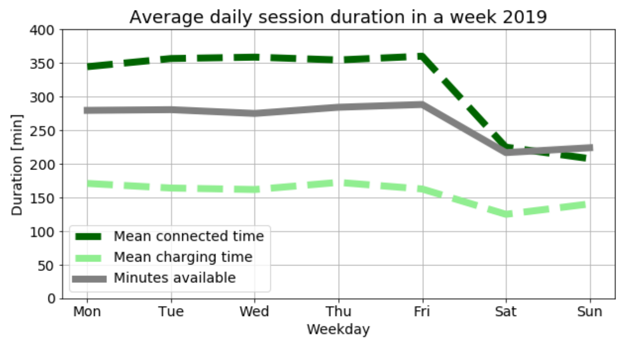
5.1.6 User-behaviour and smart scheduling

In the previous sections, the user-behaviour displayed by EV owners charging at the ACN was investigated, by solely looking at the service they were provided by the ACN. In reality, the investigated user-behaviour might be skewed by the fact that the ACN is a smart charging station, applying an ASA to schedule charging sessions interactively [22]. The duration an EV spends charging and the energy delivered in a session is heavily dependent on how the ASA schedules different charging sessions. As explained by Z. J. Lee et al., the ASA used to schedule charging sessions at the ACN uses user-input information about how long the EV owners wants to be connected, and how much energy the EV owners requests, along with measurements from the local grid to schedule the charging sessions [23].

Figure 5.8 shows the user inputs compared to the service delivered by the ACN for each day of the week, for sessions performed in 2019.



(a) Average energy delivered and user-input energy requested for a charging session each day of the week.



(b) Average session duration, charging duration and user-input minutes available for each day of the week.

Figure 5.8: User-inputs compared to the actual recorded user-behaviour at the ACN. For each day, the average value for user-inputs and average recorded user-behaviour, across all claimed sessions performed the respective day, is presented.

Figure 5.8 (a) shows the average user-input energy requested, and energy delivered for a charging session, for each day of the week. The figure shows that there was a large difference between the Energy requested by EV owners and the Energy delivered by the

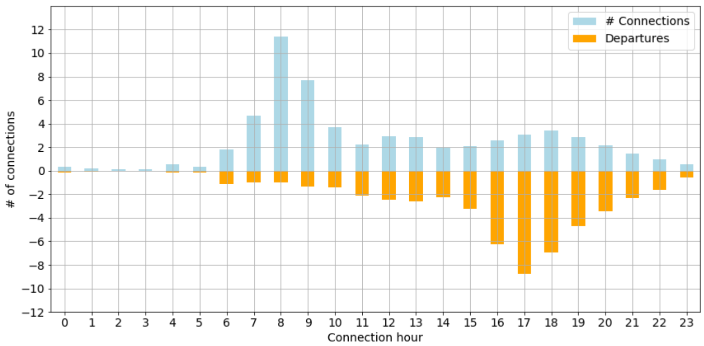
ACN. In fact, from the figure, only about half of the requested energy was delivered in an average charging session. The reason for this large difference was not explicitly described in the data, but could be a consequence of the ASA also using measurements from the grid when scheduling, and that the ASA mainly focused on not breaking grid constraints. However, as further shown in Figure 5.8 (b), the charging duration was on average shorter than the user-input Minutes available. This could suggest that, on average, the ACN was fully charging the connecting EVs before the requested deadline of delivery, provided by the EV owners. Furthermore, Figure 5.8 (b) also suggests that the EV owners charging at the ACN on weekdays, on average, left their EV connected for longer than what they stated. In addtion, on Sundays, the EV owners on average disconnected before their requested deadline of delivery. Thus, Figure 5.8, does not mainly indicate that the ASA struggled to delivered the requested energy, but rather indicates that the user-inputs given to the ASA from the EV owners were inaccurate.

With Figure 5.8 indicating that the user-inputs given by the EV owners were inaccurate, it becomes important to discuss what consequences this could have for a smart charging scheme. For example, reflecting on a DLC approach like the ASA used at Caltech. The ASA takes user-inputs from connecting EVs to schedule charging sessions for all the connected EVs. The scheduling is done in such a way that it maximises the power transfer from the station to each EV while not breaking grid constraints [23]. The main challenge contributed to erroneous user-inputs, therefore becomes that it could affect the scheduling for other EVs. If an EV owner requests a large amount of energy, the ASA could be forced to shift more of the charging load, prolonging the charging duration of other EVs. Hence, EV owners with only a limited amount of time available, might not receive their requested energy. Moreover, if EV owners stay connected for a shorter periods than what they state, it could result in EV owners not receiving their requested energy. On the other hand, if EV owners stay for longer than what they say, it could prevent other EVs from connecting and charging their EV.

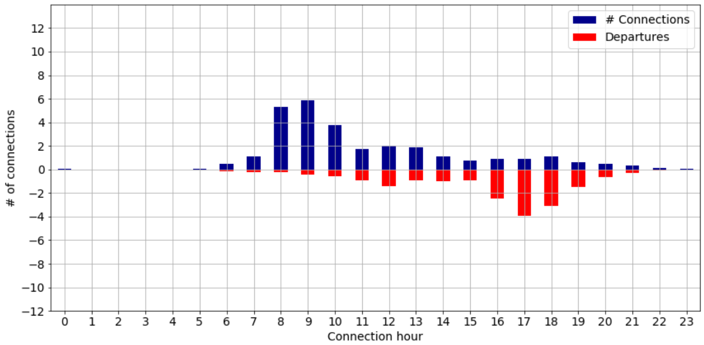
Overall, for the ACN at Caltech, the EV owners on average provide a sufficient amount of time to fully charge a majority of the EVs connecting. On average, for all days, EVs are done charging in a shorter period of time than the user-input Minutes available. On weekdays, EVs stay connected for longer than the stated minutes available which could potentially prevent other EV owners from charging at the ACN. On weekends however, EVs usually stay connected for the period stated in the Minutes available. Additionally, on Sundays, there was also a trend that EVs were connected for an even shorter amount of time. However, on average, the ASA was still fully able to schedule and fully charge EVs before the requested deadline of delivery.

5.1.7 A further look at events changing charging behaviour

Figure 5.9 displays the average daily charging pattern prior to Event 1, the 1. November 2018, and the average charging pattern for the ACN post Event 1.



(a) Average number of connections and disconnections at the charging station in 2018.



(b) Average number of connections and disconnections at the charging station in 2019

Figure 5.9: Comparison of charging pattern at the ACN before and after a charging fee of $0.12\frac{\$}{kWh}$ was implemented.

The Figure shows that after Event 1, there was a significant reduction in EVs connecting for all hours of the day. Prior to Event 1, the number of connections peaked in the morning hours with around 12 connections between 8 and 9. The number of connections then dropped to around 2 during the day. In the evening, most of the EVs connecting in the morning started to disconnect and leave. Figure 5.9 (a) then shows that the number of connections started to increase again, creating another peak in connections at around 18. However, as shown in Figure 5.9 (b) after Event 1, this changed. Firstly, there was a significant reduction in connections in the morning. The number of connections still peaked between 8 to 9, however, the the number of connections was reduced to around 6 EVs connecting. Furthermore, in the evening, when the EVs that connected in morning started to disconnect, there was no longer a significant increase in connections.

The reduction in activity at the ACN in November 2018 could be a consequence of the introduction of a charging fee at the ACN. The 1. November 2018, a fee of $0.12\frac{\$}{kWh}$ was introduced for charging at the ACN [22]. The charging fee created a cost of charging at the ACN which was similar to the price of charging at home. According to California Center of jobs and the economy, and the US Energy Agency, the residential price of electricity in California was 0.18 $\frac{\$}{kWh}$ in November 2018 [5]. Post Event 1, and the introduction of a charging fee, the charging pattern at the ACN became a more distinct workplace charging pattern. Before the charging fee was introduced, the ACN functioned as a combined workplace and public charging station. The ACN had a large morning peak as expected of a workplace charging station, but also a peak in the evening, as expected from a public charging station (see Figure 2.2). After the fee was introduced, this peak in connections in the evening disappeared. Thus, Figure 5.9 could suggest that the introduction of a charging fee led to many of the customers at the ACN, that did not work at Caltech, either chose to charge at another public charging station, or delayed charging until they arrived at home.

The second major event that had a severe impact for EV charging at the ACN was Event 2, distinguished in Figure 5.1. In March 2020, the number of EVs charging at the ACN drastically reduced, to where there were almost no EVs charging at the ACN. The impact of Event 2 can be seen for the rest of 2020. Figure 5.10 shows the average number of EVs connecting each hour in a day post Event 2.

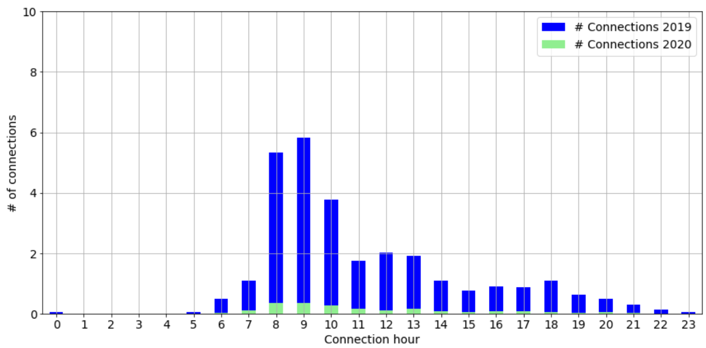


Figure 5.10: Comparison of charging pattern at the ACN before and after the COVID-19 pandemic caused a statewide lock down in California.

As shown in the figure, in 2020 the average number of connections each hour in a day dropped to almost zero. The peak of activity was still in the morning between 7 and 9, however, whereas in 2019 the average number of connection peaked at almost 6 per hour, in 2020 it was less than 1 EV connecting.

Event 2 could display the impact the COVID-19 pandemic had on the ACN, and the EV owners charging at the ACN. The 19. of March 2020, Gavin Newsom, the Governor of California issued an executive order that prohibited large gatherings and called for California's residents to stay and conduct their work at home [30]. The executive order led to Caltech closing down all in-person activity, except for essential work [38]. For the ACN this meant that workers that usually charged their EVs during weekdays no longer drove to Caltech, and instead charged at home, or closer to home. In addition, EV owners who used the ACN as a public charging station, no longer drove to Caltech as all residents were encouraged to stay at home [30].

5.2 Analysing user-behaviour in charging groups

After the general charging pattern for the ACN and patterns in the average user-behaviour was investigated, the focus shifted to analysing if the EV owners could be categorised into distinct charging groups. The charging groups could then further be used to investigate if EV owners displayed different user-behaviour characteristics accordingly. As mentioned, the ACN dataset contains historical charging data, and for some of the recorded charging sessions, unique user ids and user-inputs were included (claimed sessions). Using these claimed sessions, it was possible to investigate the frequency EV owners charged at the ACN, thus enabling the possibility to categorise EV owners into charging groups based on how often the individual EV owners charged at the ACN.

The number of claimed sessions recorded in the dataset was, however, significantly lower than the total number of sessions. Hence, to create and analyse charging groups that were representative of all charging at the ACN, it was important to identify a period where the number of claimed sessions was similar to the total number of sessions. Figure 5.11 shows the distribution of all sessions and claimed sessions for the entire downloaded period.

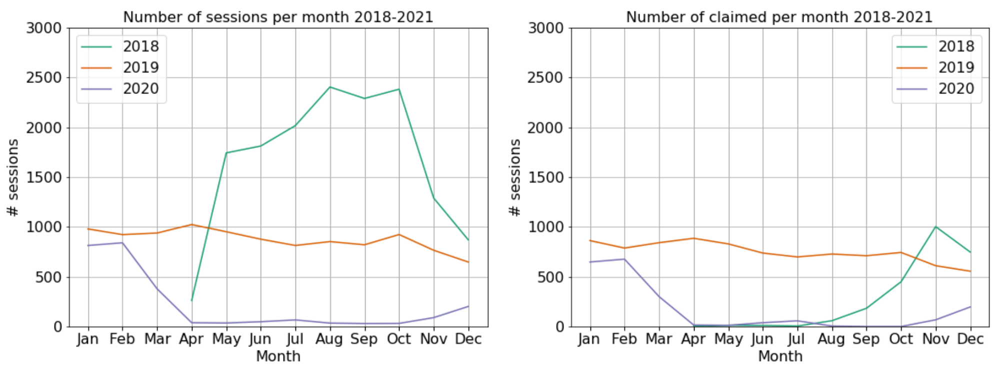


Figure 5.11: Distribution of charging sessions and claimed charging sessions for the period 25. April 2018 to 1. January 2021.

The figure shows that only a small part of the recorded charging sessions in 2018, prior to the inception of the charging fee, were claimed sessions. The reason for this was that the ACN only started recording user-inputs from EV owners in August 2018 [22]. In 2019 and 2020, most of the sessions were claimed sessions. Thus, for 2019 and 2020, it could be assumed that claimed sessions could be used to analyse EV user-behaviour representative of all EVs charging at the ACN. Unfortunately, due to the effects of the COVID-19 pandemic, there was overall very few sessions recorded in 2020. This had a significant impact on the frequency of which EV owners charged. Some EV owners charged very frequently at the ACN in 2019, distinctly more than others. The COVID-19 pandemic reduced this frequency, as almost no EV owners frequently connected to the ACN in 2020. Hence, blurring the boundary between charging groups. Therefore, for investigating differences in user-behaviour across charging groups, only sessions from 2019 were used. In 2019, most sessions were recorded as claimed sessions and the overall charging pattern, and user-behaviour was not impacted by any major external event, at least none seen in the dataset.

5.2.1 Defining the charging groups

Figure 5.12 shows a distribution of the the number of charging sessions recorded for each unique user id.

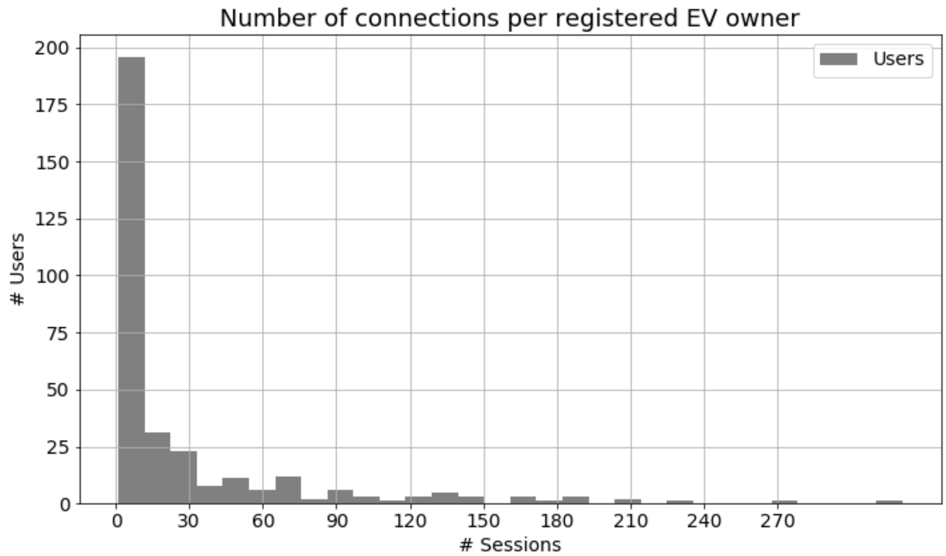


Figure 5.12: Distribution of number of charging sessions per registered EV owner charging at the ACN in 2019.

Figure 5.12 shows that most registered EV owners charged less than 50 times at the ACN during 2019. This means, on average less than once a week. Furthermore, the figure also show that there was a significant amount of EV owners who used the ACN more frequently.

The EV owners were separated into three distinct charging groups. These groups were Sparse users, charging less than once a week (<52), Weekly users, charging on a weekly basis (>52, <104), and Daily users, who charged more than two times a week (>104).

Table 5.3 gives the number of EV owners falling into each charging group, and the number of charging sessions performed by each charging group.

Table 5.3: Number of EV owners and number of charging sessions performed by the EV owners, in each charging group.

EV Charging groups

	All	*Sparse*	*Weekly*	*Daily*
# of EV users	322	264	34	24
# Connections	8984	2473	2457	4054

The general idea of the charging groups was to capture the differences in displayed user-behaviour of EV owners who used the ACN inherently different. The reasoning behind Daily users was to capture trends in EV user-behaviour of EV owners who frequented Caltech, like students and employees. With the Weekly users, the idea was to capture trends in EV user-behaviour of EV owners who were often in the vicinity of Caltech, for example, guest lecturers and EV owners who lived nearby. The idea behind Sparse users was to capture the EV user-behaviour of EV owners who usually did not visit Caltech, and might have only stopped to charge to be able to reach their final destination. Figure 5.13 further shows the distribution of number of sessions per EV owner categorised in each charging group.

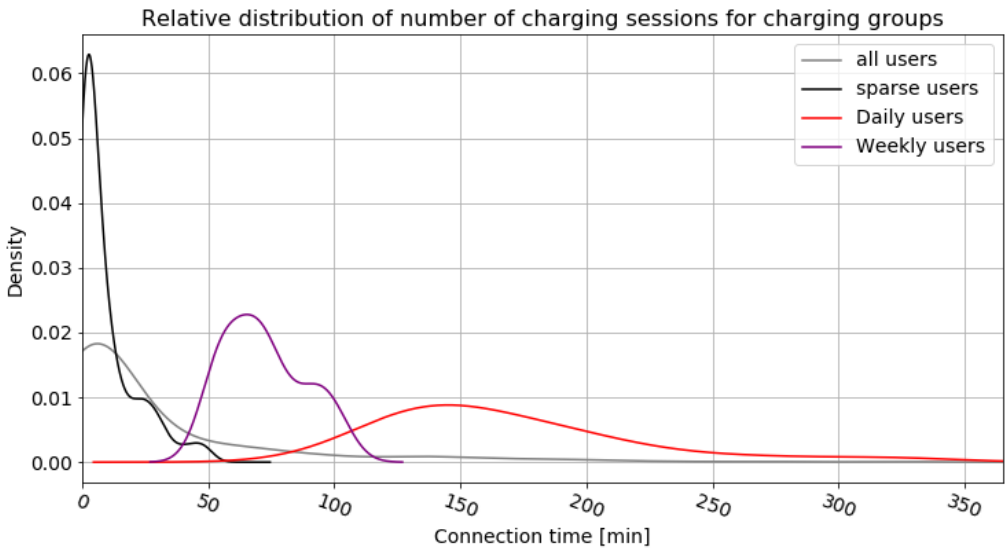


Figure 5.13: Relative distribution of number of charging sessions performed by each EV owner in each charging group.

These three distinct charging groups were further used to investigate whether EV owners categorised into each group, displayed a distinctly different EV user-behaviour. This was analysed by investigating the times they arrived, how long they usually left their EV connected, and how much energy they charged their EV. Lastly, the charging groups were also used to investigate if EV owners in each group, overall, provided a different accuracy for their user-inputs.

5.2.2 Average connection time for the charging groups

Figure 5.14 shows the relative distribution of the average connection time for each EV owner in each charging group.

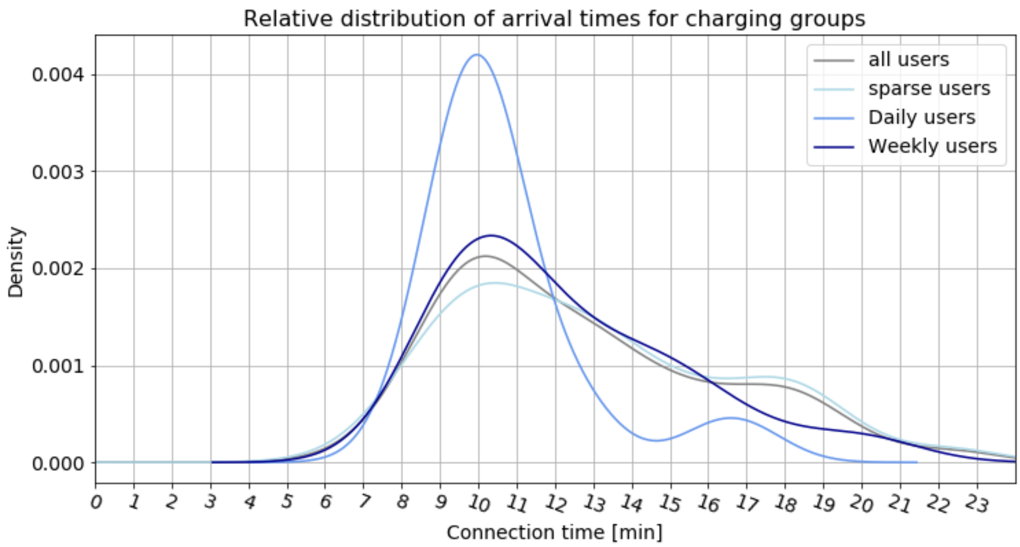


Figure 5.14: Relative distribution of average connection time for the EV owners in each charging group.

The figure shows that for all charging groups, and all EV owners in total, most EV owners on average connected their EV to the ACN in the morning between 9 and 11. However, this was most significant for the Daily users. For Weekly users, the distribution of arrival times was a little more spread out, with a more rounded peak in the morning. For Sparse users, there was less of a significant peak between 9 and 11. The connections were overall more distributed throughout the day.

Overall, with the distribution of the connection times, it was possible to see how the EV owners in the different charging groups on average used the ACN. The distribution for the Daily users suggested that most of these Daily users utilised the ACN as a workplace charging station, connecting their EV when arriving at work in the morning. However, the distribution also identified that a small portion of the Daily users that on average connected later in the day between 16 and 17. This could suggest that within the Daily users charging group, there were at least two groups of EV owners displaying dissimilar behaviour when it came to connection time. For Sparse users, the distribution displayed that Sparse users connected more spread out throughout the day, and the observed distribution was similar to that expected for a public charging station, as seen in Figure 2.2. The Weekly users were somewhere in between, with a large portion of the Weekly users connecting in the morning, but still a substantial share connecting later in the day. This could suggest that the connection time might not be a defining characteristic trend for EV owners categorised as Sparse users and Weekly users.

5.2.3 Average charging session duration for charging groups

Figure 5.15 shows the relative distribution of the average session duration for each EV owner in each charging group.

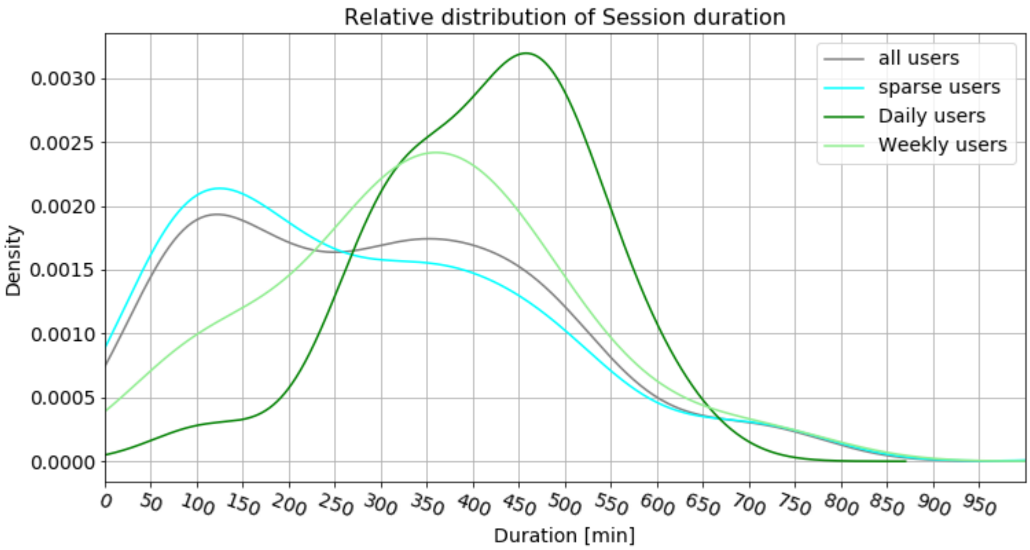


Figure 5.15: Relative distribution of the average Session duration for the EV owners in each charging group.

The figure shows that the average duration each EV owner left their EV connected, varied between the charging groups. In Figure 5.15, the distribution of the measured session duration is shown. As visualised, the Daily users tended to leave their EV connected for longer than the other groups. For Daily users there were two distinct peaks in the distribution. One for Daily users leaving their EV connected between 300 minutes and 350 minutes, and a larger one for Daily users leaving their EV connected between 450 minutes and 500 minutes. For Weekly users, there were also two peaks in the distribution, a smaller one between 100 minutes and 150 minutes, and larger peak between 320 minutes and 370 minutes. For Sparse users, most of the EV owners on average left their EV connected between 100 minutes and 150 minutes. However, the distribution for Sparse users also displayed that a significant amount of Sparse Users left their EV connected for longer, and some even left their EV connected for 700 minutes or more.

Figure 5.16 shows the relative distribution of the average user-input minutes available for each EV owner in each charging group.

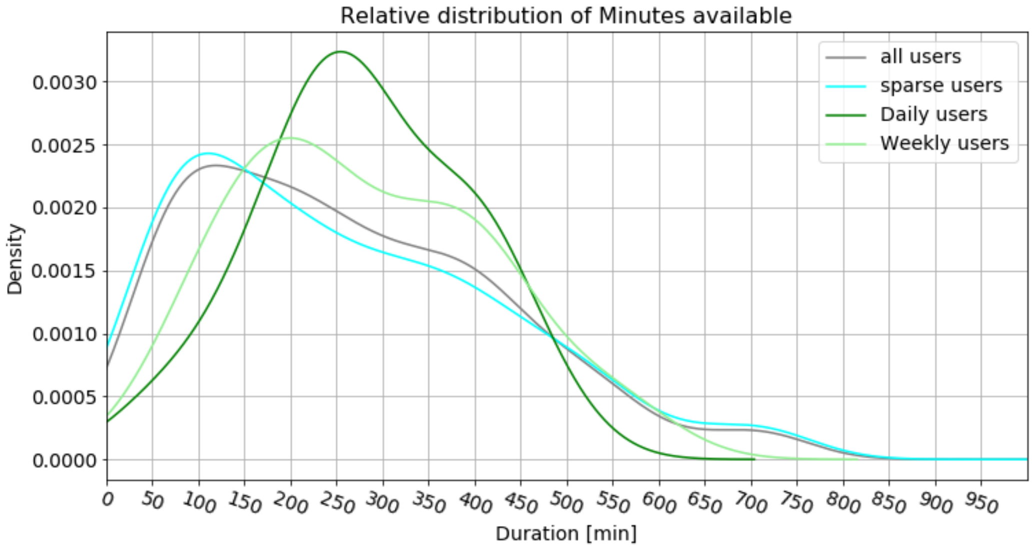


Figure 5.16: Relative distribution of the average user-input minutes available from the EV owners in each charging group.

The distribution for minutes available seen in Figure 5.16, varied significantly from the distribution of the actual recorded session duration (see Figure 5.15). From the distribution, it was found that Daily users tended to provide a shorter duration than what they actually left their EV connected. The minutes available provided by Sparse users on the other hand, tended to be a longer duration than what they actually left their EV connected. For the Weekly users, the distribution shows that most of the Weekly users, either left at a similar time to what they stated in the user-inputs, or disconnected prior to their stated departure time.

Overall, Figure 5.15 and Figure 5.16 shows that the distributions for Sparse and Weekly users have two peaks around significantly different durations, both for the session duration, and the minutes available. This could suggest that within these two charging groups, there were groups of EV owners displaying a distinctly different user-behaviour at the ACN. This could be the difference between EV owners who are able to charge their EV at home, and those who are not. For the Daily users there were also two peaks in the distributions. These peaks could be a consequence of, for example, Daily users owning EVs with different battery packs, or Daily users who attend Caltech and have different work schedules.

5.2.4 Average energy delivered for the charging groups

Figure 5.17 shows the relative distribution of the average energy requested for each EV owner in each charging group.

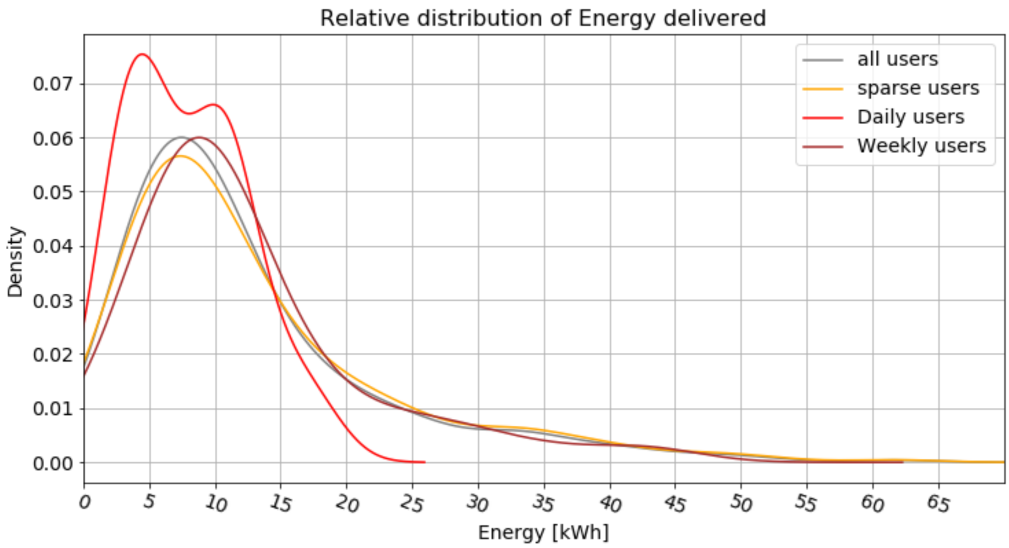


Figure 5.17: Relative distribution of the average Energy delivered to the EV owners in each charging group.

Figure 5.17 shows that for daily users there were two distinct peaks for the average energy delivered, per charging session. One peak around 5 kWh and one around 10 kWh. For the Sparse and Weekly users however, there was a wider peak in the distribution ranging from around 5 kWh to around 10 kWh. The distribution also identified a small share of the Sparse and Weekly users that on average received significantly more energy, receiving between 40 kWh and 45 kWh on average per charging session.

Figure 5.17 displays that there were two peaks in the distribution for daily users. The larger of these peaks indicated that more of the Daily users requested less energy than the majority of the EV owners in the other groups. This could be a consequence of that these EV owners charged their EV often, and due to their driving habits therefore required less energy to charge their EV each time they charged. The smaller peak, around 10 kWh delivered on average per charging session, indicate that a significant share of Daily users charged their EV with a similar and slightly larger amount of energy than the EV owners in the other charging groups. This could be a consequence of the distance these EV owners had to commute to arrive at the ACN or driving could be a part of their job.

Figure 5.18 shows the relative distribution of the average user-input energy requested by EV owners in each charging group.

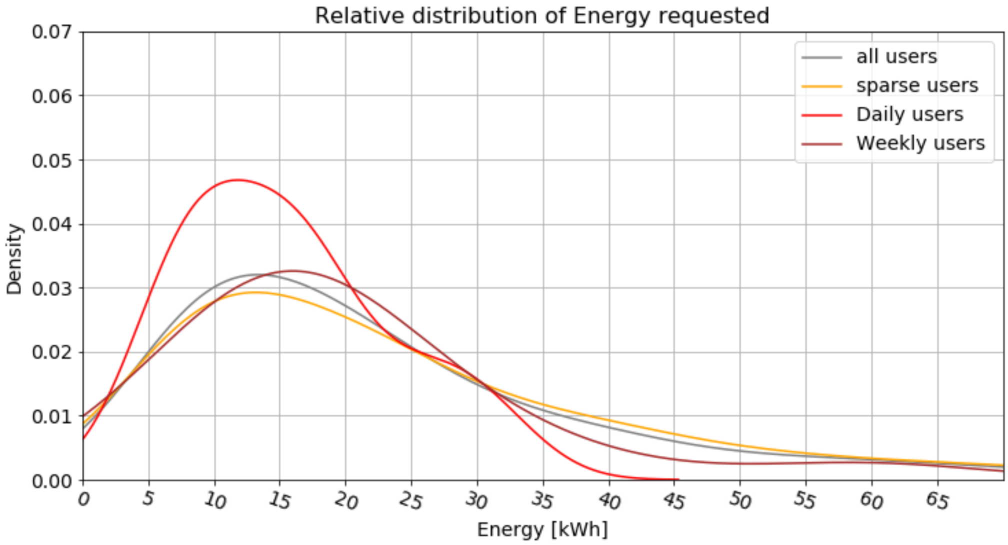


Figure 5.18: Relative distribution of the user-input energy requested from the EV owners in each charging group.

From the distributions in Figure 5.18 and Figure 5.17, it was found that all EV owners, on average asked for more energy than what they received. For all groups, most EV owners requested between 10 kWh and 15 kWh as seen in Figure 5.18. However, there was also a significant amount of EV owners in each of the charging group that asked for more than this. For both Weekly and Sparse users, there were multiple EV owners that on average, asked for upwards of 55 kWh to 60 kWh. The Weekly users that requested a high amount of energy, did this on a regular basis, as all of the Weekly users had charged more than 52 times in 2019. The Sparse user distribution however, was based on much fewer observations per EV owner, thus the mean value used to represent a Sparse user were more exposed to outlier values.

Overall, Figure 5.17 and Figure 5.18 did not identify an inherently different behaviour when it came to the energy demand of the EV owners. The distributions for the Daily users were more concentrated than for the other charging groups, but the majority of EV owners in each group both asked for and received a similar amount of energy, compared to the resolution of the distribution. As the distributions were made to capture every EV owner, the distributions became general and skewed due to some EV owners both received and requested very large amounts of energy compared to other EV owners. It is possible that if some of these EV owners were filtered out in the pre-processing, it could have been easier to identify and describe more differences.

5.2.5 Error in user-inputs for the charging groups

Figure 5.19 show the relative distribution in MAE for the user-input minutes available for each EV owner in each charging group.

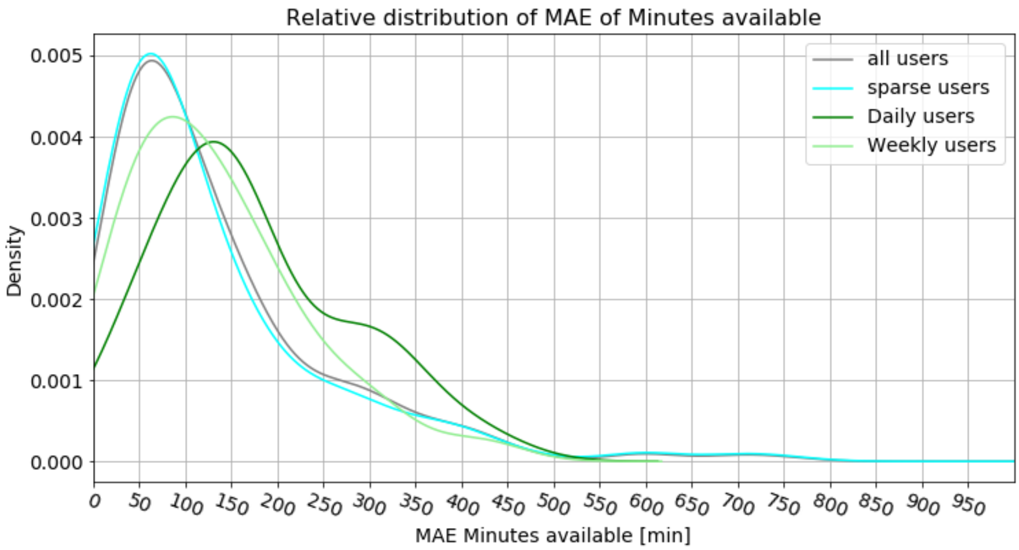


Figure 5.19: Distribution of MAE in user-input minutes available from the EV owners in each charging group.

From the distribution in Figure 5.19, it was found that the Daily users overall provided the highest MAE of all the charging groups. For the Daily users, the distribution peaked at a MAE around 120 minutes to 150 minutes, however, some Daily users also provided a Minutes available with a higher MAE, as identified by a smaller peak at around 270 minutes to 310 minutes. On the other hand, the charging group providing the most accurate overall user-inputs, with the lowest MAE, were Sparse users.

It is however important to note that the error provided in the user-input minutes available from the Daily users, was mainly due to the Daily users leaving their EV connected for longer than stated. For the Sparse users, the error was mainly due to Sparse users disconnecting before the stated departure time in the user-input minutes available. Thus, for scheduling the charging sessions, the error in the Sparse users' user-inputs could lead to Sparse users leaving before they received their requested energy. The Daily users on the other hand, could prevent other EV owners from charging at the ACN, as Daily users usually left their EV connected for longer than what they stated.

Figure 5.20 shows the relative distribution in MAE of the user-input energy requested for the EV owners in each charging group.

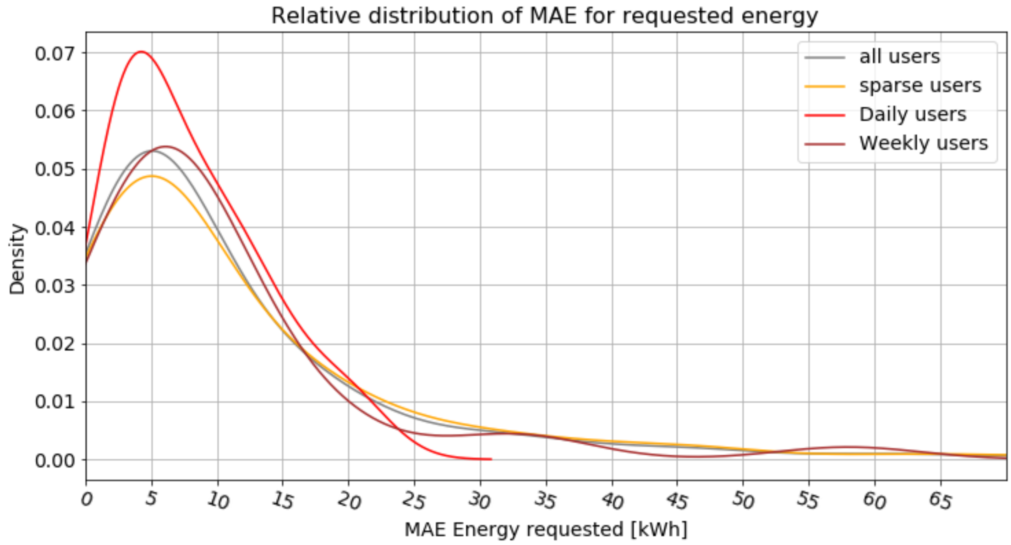


Figure 5.20: Relative distribution of MAE in user-input energy requested from the EV owners in each charging group.

Figure 5.20 shows that charging groups had a similar distribution of MAE for the user-inputs from the EV owners, with respect to the requested energy. The distribution for MAE for all three groups peaked at around an MAE of 4 kWh to 6 kWh. It was observed in Figure 5.17 and Figure 5.18 that this error in the requested energy was mainly due to EV owners requesting more energy than what they received.

Overall, the analysis of variation in energy demand between charging groups was weakened by EV owners requesting and receiving large amounts of energy. However, the error in the user-inputs could still provide some insight into the EV user-behaviour seen at the ACN. It was observed that the most EV owners generally requested more energy than they received both in Figure 5.8 and Figure 5.20. Moreover, Figure 5.8 suggested that EVs charging at the ACN was on average fully charged before they disconnected. Additionally, in the pre-processing, only 1152 samples from the entire sample pool where identified to not have a Done charging time. This could suggest that the EV owners that charged at the ACN did not understand how much energy was needed to charge their EV in terms of kWh, which they were requested to provide. Hence, most EV owners requested a large amount of energy, as a way to ensure that their EV was fully charged. It is possible that encouraging the EV owners to input a measure of how much energy they want in terms of something they can more easily relate to, like what is displayed on the dashboard in their EV, could improve the accuracy of the user-inputs. E. g. it could be interesting to investigate if asking the EV owners for a requested state-of-charge instead of a specific amount of energy could lead to a higher accuracy.

5.2.6 Weaknesses and improvements for the charging groups

Overall, the defined charging groups gave some insight into how EV user-behaviour varied for EV owners that charged at the ACN. However, as was found when investigating the relative distributions, the charging groups were too general to accurately separate EV owners. E.g. when investigating the connection time, at least two groups were found within the Daily users group, and no distinct patterns were found for Sparse and Weekly users. Likewise, when investigating the session duration, at least two groups were identified within each of the charging groups. This could suggest that purely separating the EV owners based on the number of charging sessions they performed, did not accurately distinguish the charging groups. Additionally, the resolution of the distribution for energy requested became too general to identify differences between the charging groups. A reason for this could have been that it was decided to not filter out charging sessions with abnormally large values in the Energy requested feature. Another weakness of the analysis of charging groups was that each EV owner was only represented by their average behaviour based on all their charging sessions. For this to be accurate, it relies on EV owners charging their EV in a distinct and repetitive manner across all sessions. In reality, this might not be true. It could be interesting and beneficial to investigate how diverse the charging sessions of EV owners are. Lastly, the Sparse users charging group consisted of a majority of EV owners that only charged 2 to 4 times during the entirety of 2019. Attempting to analyse the user-behaviour of an EV owner with such an inadequate amount of data is unlikely to represent the general charging behaviour of the EV owners.

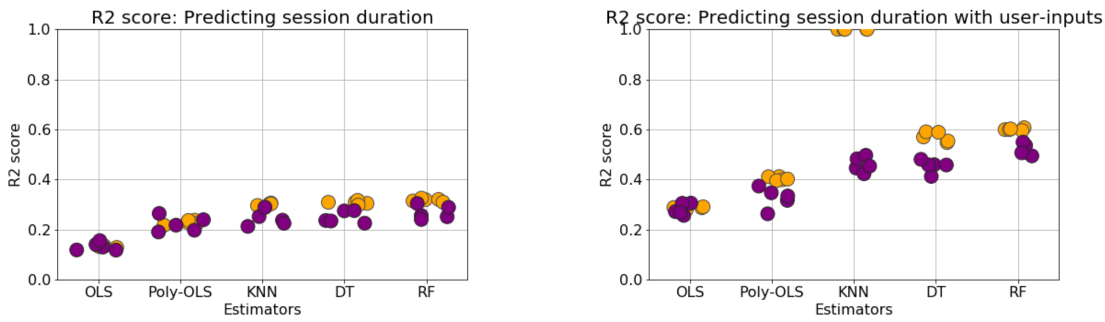
5.3 Predicting EV user-behaviour with Machine Learning

Due to the increasing focus of utilising Machine Learning and Big data analytics, for solving the EV charging problem, and for modelling the EV charging load, it was attempted to use the data available in the ACN dataset to predict some aspects of EV user-behaviour. Machine Learning models were built for predicting the expected session duration and the expected energy delivered in a charging session, using two different subsets of the ACN dataset.

The results from the Machine Learning was compared with the user-inputs to investigate which gave the most accurate depiction of the observed session duration and energy delivered. Finally, the most reliable model for predicting session duration, and the most reliable model for predicting energy delivered were used to identify features that had the most significance for the performance of the models.

5.3.1 Predicting Session duration

Five Machine Learning models were trained for predicting session duration, using two different subsets of the ACN dataset. One subset only providing the connection time and the weekday dummy features, and one providing some additional information about the EV connecting. This additional information consists of the User id feature and the Energy requested feature. Figure 5.21 shows the overall performance when training and predicting on new data, represented with an R2-score.



(a) Performance of models with only connection times as input.

(b) Performance of models with connection times, user id and Energy requested as input.

Figure 5.21: Performance of the supervised learning regression models predicting session duration. The figure displays the R2-score for training and predicting for a 5-fold cross-validation for each model.

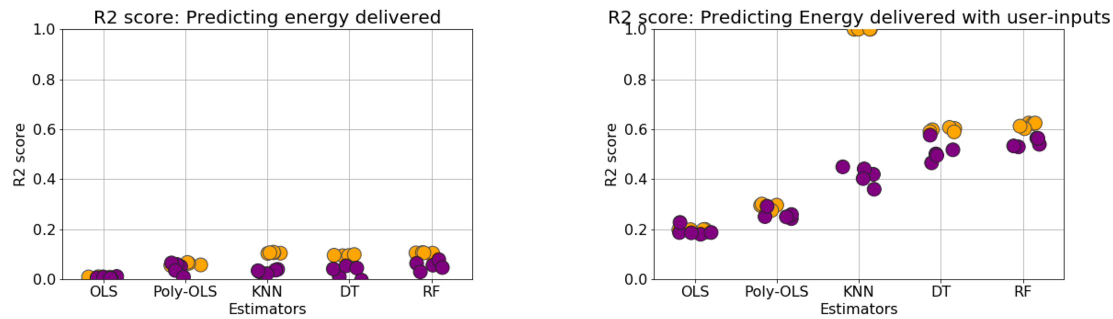
Figure 5.21 displays the R2-score for training (orange) and predicting (purple) achieved in a 5-fold cross-validation for each model. Figure 5.21 (a) visualises the performance of the models when trained and tuned, using only information about the connection times. Overall, the models had a similar performance for training and predicting on new data, achieving a poor R2-score for both. The KNN model, the DT model and the RF model had superior performance, achieving an R2-score of around 0.30. Figure 5.21 (b) visualises the performance of the models when trained on a larger subset of the ACN dataset. This larger subset providing more information about the EV owner connecting and the requested energy, led to a significantly better R2-score for all models. The RF model gave the highest performance overall, with a R2-score of around than 0.50 when predicting on new data. This was significantly better than when trained on the smaller subset.

Figure 5.21 also indicate how well each of the models generalised when introduced to new data. Figure 5.21 (a) show that none of the models were suffering from overfitting when trained on only the connection times. For all models, the training and prediction scores were similar for all cross-validation folds. This meant that the models were able to apply the found relationship to predict the target variables when input new data, with a similar performance to when training. However, the reason for this might be that none of the models found a clear relationship, as the overall performance was quite poor. In Figure 5.21 (b) however, overfitting became more prevalent with the larger subset. Except for the OLS model, all the other models displayed a tendency to overfit, with a higher R2-score for training than when predicting. The clearest overfitting model was the KNN model. On the smaller subset, with only connection times, the KNN model had a similar R2-score for training and predictions, but when trained on the larger subset with more information, the model overfitted heavily. As a consequence of the overfitting, the R2-score for the KNN model for training was 1.0, while for predictions it was around 0.50.

It was attempted to reduce the overfitting of the KNN model by increasing the number of neighbours used to predict, but with no success. The R2-score for training remained at 1.0, even with 'k neighbours' increased to a high number, like 100, 200 and even 500. The only observed impact was a reduced R2-score for predictions.

5.3.2 Predicting Energy delivered

Five machine learning models were trained for predicting the expected energy delivered for a charging session, using two different subsets of the ACN dataset. One subset providing only the connection time and the weekday dummy features. The other subset providing some additional information about the EV owner connecting, namely, the User id and Minutes available features. Figure 5.22 shows the overall performance when training and predicting on new data, represented by a R2-score.



(a) Performance of models with only connection times as input.

(b) Performance of models with connection times, user id and Minutes available as input.

Figure 5.22: Performance of the supervised learning regression models predicting Energy delivered. The figure displays the R2-score for training and predicting for a 5-fold cross-validation for each model.

Figure 5.21 (a) visualises the performance of the models when trained and tuned using only information about the connection times. Overall, the models, similarly to the models predicting session duration, gave poor performance, achieving a low R2-score. The models were even more limited in their capacity to model a relationship between the input data and target variable 'Energy delivered', than the target variable 'Session duration'. This was indicated by all models achieving a R2-score lower than 0.1, across all five folds of the cross-validation, both for training and predicting. None of the models displayed any significantly better performance than the others, and they all struggled to model a relationship, which was reflected by the poor R2-scores. Figure 5.21 (b) visualises the performance of the models when trained on the larger subset of the ACN dataset. This larger subset providing the User id feature and the Minutes available feature, led to a significantly improved R2-score for all models. As can be seen from Figure 5.21, all models achieved a greater R2-score, but more significantly, some of the models went from not modelling a accurate relationship at all, to achieving R2-scores higher than 0.50. The best performing models were the RF model and the DT model.

As seen from Figure 5.21 (b) training the models on a larger subset of data, made some of the models overfit, most distinctly the KNN model. During training the KNN model achieved a R2-score of 1.0. Seemingly modelling a perfect relationship between the input data and the target variables. However, when introduced to new data, the R2-score was only around 0.40. This suggested that this KNN model was also severely overfitting when used with the larger subset. Tuning the KNN model did not reduce the overfitting, however it reduced the R2-score when new data was input. The fact that the KNN models overfitted for both cases where the larger subset was used, could suggest that some of the features that was included in the larger subset could have a high correlation with the target variables.

5.3.3 Comparison of predictions and user-inputs

Overall, the Machine Learning models achieved a limited R2-score, and were unable to produce accurate predictions for session duration and energy delivered. This was however also the case for the user-inputs provided by the EV owners at the ACN. As found both when investigating charging groups and investigating user-behaviour at a general station level, user-inputs where also quite inaccurate compared to the service delivered by the ACN. Therefore, it was investigated if the accuracy of the predictions from the best performing Machine Learning models, either rivalled the accuracy of the user-inputs, or actually gave better estimates.

The best performing Machine Learning models for predicting the session duration and the energy delivered were both found to be RF models. Table 5.4 shows each of the RF models with the parameters found in the grid search.

Table 5.4: Overview of the best performing models with the tuned parameters found in the grid search.

Tuned RF models:

RF model for predicting Session duration					
Parameter	*Max depth*	*Min sample split*	*Min sample leaf*	*Criterion*	*# estimators*
Value	9	3	10	MSE	100
RF model for predicting Energy delivered					
Parameter	*Max depth*	*Min sample split*	*Min sample leaf*	*Criterion*	*# estimators*
Value	9	2	10	MSE	100

The MAE for the predictions for the models presented in Table 5.4, are further given in Table 5.5, along with the MAE for the user-inputs.

Table 5.5: MAE of predictions from the supervised regression models and the user-inputs.

Comparison of predictions and user-inputs

	Predictions	*User-input*
MAE Session Duration [min]	104	149
MAE Energy delivered [kWh]	4.0	9.2

In Table 5.5, the MAE for the user-inputs is the difference between the minutes available feature and the observed session duration feature, and the difference between the energy requested feature and the observed energy delivered feature. Although the MAE for predictions was quite large, with an MAE of 104 minutes for session duration and 4.0 kWh for energy delivered, the MAE for predictions were significantly lower than for the user-inputs. Compared to the observed session duration and energy delivered, the user-inputs had an MAE of 149 minutes and 9.2 kWh for minutes available and energy requested, respectively.

The RF models were actually able to produce more accurate estimates of the session duration and the energy delivered, than the user-inputs, when using the data available in the ACN dataset. This indicates the potential for applying Machine learning to the smart charging scheme utilised at the ACN. By improving the models further, they could potentially be to used for smart charging accompanied by a smart scheduling algorithm.

5.3.4 Identifying important factors for user-behaviour

The final investigation of the Machine Learning models was focused on identifying features that were important for the performance for the models. This could be interesting for station operators or other researchers who desire to build datasets to analyse and model EV user-behaviour further.

To identify important features from the Machine Learning, only the the RF models were investigated as they achieved the highest R2 score for both predicting session duration and energy requested. With the RF model provided in the Sci-kit learn library, it was possible to retrieve feature importances that displayed how important each feature was for the performance of the model. The calculated feature importances sums to one, and the importances are distributed to the features based on the average accumulated impurity decrease each feature is responsible of in the DTs in the RF model [36]. It is however important to note that due to the limited R2-score achieved for predictions with the RF models, the feature importances that were found could be inaccurate. This is because the R2-score describes the amount of variance in the target variable the model is able to find with the variation in the provided features. As the R2-score was only about 0.50 for each model, the models were only able to find a relationship describing about half of the variation in the target variables. Thus indicating that there could be other features that are of more significance for explaining the true target variables.

Figure 5.23 shows the feature importances for predicting the Session duration with the tuned RF model.

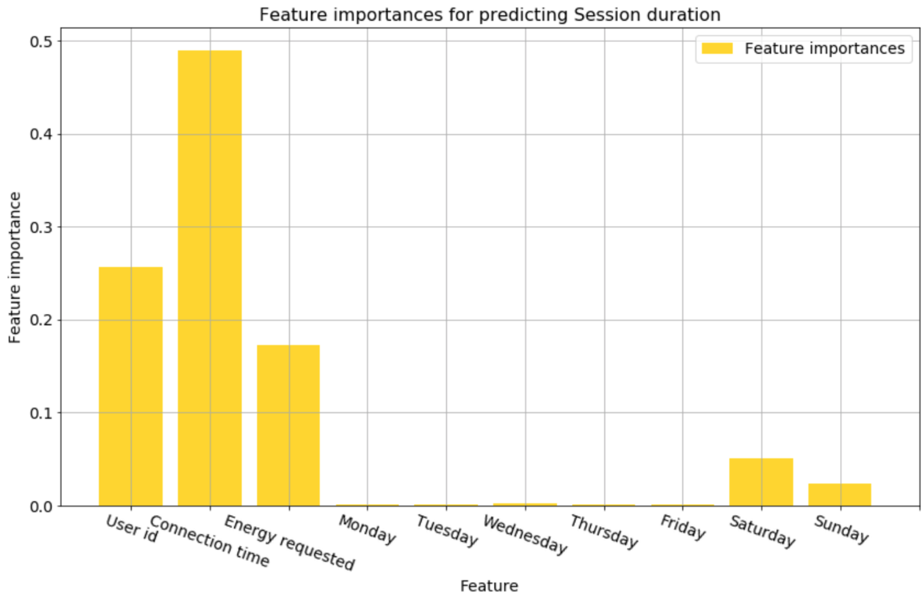


Figure 5.23: Feature importances for the performance of the Random Forest regression model predicting session duration.

The figure shows that the most important features for the performance of the RF model, was 'Connection time', 'User id' and 'Energy requested'. The most important feature was 'Connection time', with a calculated importance of around 0.48, followed by the 'User id' with an importance of around 0.25, and 'Energy requested' with an importance of around 0.18. In addition to this, some importance was attributed to the Saturday and Sunday features.

Figure 5.24 gives the feature importances for predicting energy delivered with the tuned RF model.

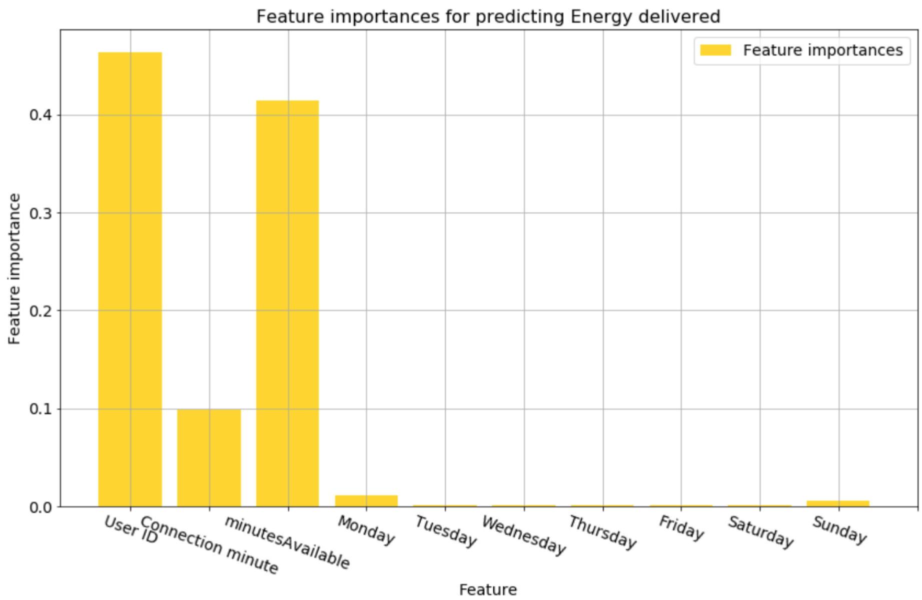


Figure 5.24: Feature importances for the performance of the Random forest regression model predicting energy delivered.

The figure shows that the most important features were the 'User id', the 'Minutes available' and the 'Connection time'. The 'User id' was attributed the greatest feature importance with around 0.48, followed by 'Minutes Available' with around 0.41, and 'Connection time' with around 0.10. For predicting energy delivered, the Weekday features were observed to have little importance.

Overall, the feature importances were coherent with the other discoveries in the other parts of the analysis. For the connection time, it was found when investigating the general charging pattern across all charging sessions, that the duration an EV was connected to the ACN was influenced by when the EV connected. In Figure 5.5 it was observed that EVs connecting in the morning hours between 8 and 10 usually left their EV connected for longer than the EV owners connecting later in the day. The RF model predicting session duration also found that weekends had some significance. This could be a consequence of the observations in Figure 5.6. Figure 5.6 indicate that, on average, EV owners left their EV connected for longer on weekdays than on weekends. Similarly, none of the features giving weekdays showed any significant importance for predicting energy delivered could be reflected in Figure 5.7. Figure 5.7 did not indicate that there was a significant dependency for weekdays or weekends on how much energy an EV received during a charging session. Furthermore, the User id was found to have a relatively large importance for both predicting session duration and energy delivered. When investigating charging groups, it was found that EV owners charging at different frequencies at the ACN displayed different behaviour when it came to connection time and session duration. This could be a reason for the importance of the User id feature, as during training, the RF model could capture the pattern of distinct groups of users that either display similar or dissimilar user-behaviour.

Finally, to summarise the factors that were found to display the highest impact and descriptive value of the EV user-behaviour observed at the ACN. The observations in this thesis, using the ACN dataset, suggests that the most valuable information for explaining EV user-behaviour were captured in 3 distinct features. The 'User id' feature, which made it possible to separate between the EV owners charging at the charging station. The 'Connection time' and 'Disconnect time' features, which made it possible to track how long EV owners left the EV connected. Additionally, when investigating Event 1 in Figure 5.9, it was observed that the introduction of paid charging could have had an impact on the EV owners charging at the ACN. Thus, suggesting that logging a price for charging for each charging session could also be of interest.

6. Conclusion

To conclude, in this thesis, it was conducted a data search to create an overview of some open source EV charging datasets. In addition, a data-driven analysis of EV user-behaviour was conducted using one of the datasets identified in the data search.

The data search identified four sources of charging data from four different countries. These data sources were the ACN-dataset from Caltech in California [22] , the Elaad dataset provided by Elaad.NL in the Netherlands [11], the Ireland and Northern Ireland dataset provided by J. Burkin [4], and the NOBIL database in Norway [21].

The first part of the analysis investigated the charging station at an aggregated level, trying to identify significant events that had a long-term impact on the charging activity at the ACN, a charging pattern, and general trends from the observed EV user-behaviour. Two major events that had a long-term impact on the ACN, were identified. One event in November 2018 and one in March 2020. These events could be consequences of the implementation of a $0.12\frac{\$}{kWh}$ charging fee the 1. November 2018, and the impact of the lockdown caused by the COVID-19 pandemic 19. March 2020. The identified charging pattern at the ACN was based on an average across all charging sessions conducted in 2019. It was observed that the charging pattern at the ACN had a significant weekday-weekend variation, and was mostly representative of the charging pattern expected for a workplace charging station. Furthermore, investigating the aspects of EV user-behaviour on a general level, found that the EV owners charging at the ACN mostly left their EV connected for significantly longer than the period it took to charge their EV. Further investigating the user-inputs also found that EV owners tended to leave their EV connected for longer than their requested connected time. When investigating the energy demand at the ACN, it was observed that the EV owners in general asked for more energy than the ACN was able to provide. This could be due to the ACN not being able to provide the desired service. However, based on the observations, the discrepancy between the requested energy and the delivered energy was most likely due to EV owners requesting more energy than their EVs could receive.

In the second part of the analysis, it was attempted to categorise the EV owners at the ACN into three distinct charging groups. The charging groups were based solely on the frequency of which the EV owners charged. It was found that the EV owners categorised into the three charging groups displayed a different behaviour when in it came to time of connection and how long they left their EV connected. The most frequent users, the Daily users, usually connected in the morning between 9 and 11, while the EV owners in the other charging groups usually connected more spread out through the day. It was also found that the Daily users tended to leave their EV connected for longer than the other groups. The charging groups were, however, found to be too general and not able to separate the EV owners accurately. For both Connection time, and Session duration, multiple groups of EV owners were found within each charging group and when investigating differences in energy demand between the charging groups, the analysis was unable to identify any major differences.

In the third part of the analysis, it was attempted to predict aspects of EV user-behaviour with Machine Learning. Namely, the duration of a charging session and the energy delivered in a charging session. Overall, the Machine Learning models were unable to give accurate predictions for both charging session duration and energy delivered. The best models for both cases were found to be RF-models each achieving a R2-score of around 0.50. However, compared to the accuracy of the user-inputs, the RF-models' predictions were more accurate. For Session duration the user-inputs had an MAE of 149 min, while the predictions of the respective RF model had an MAE of 104 min. For Energy delivered, the user-inputs had an MAE of 9.2 kWh, while the predictions of the respective RF had an MAE of 4.0 kWh.

With the observations from all three parts the analysis, factors that could be interesting to record at a charging station, for the purpose of creating datasets to analyse EV user-behaviour, were proposed. These factors were the connection and disconnection time, marking the duration of a charging session, an id to separate between the EV owners performing the charging sessions, and the price of charging for each charging session.

7. Further work

Overall, this thesis provided a data-driven analysis investigating some aspects of EV user-behaviour at a charging station. However, there were several aspects of the analysis that could be improved or elaborated as part of a further work. Firstly, the charging groups that were used in this analysis were found to be limited in their capacity to separate the EV owners. It could be possible to apply unsupervised learning and clustering to improve the charging groups. With clustering, it would be possible to create more advanced charging groups which could enable a more detailed and accurate separation of EV owners. Secondly, the machine learning in this thesis attempted to predict the recorded Session duration and Energy delivered for charging sessions. To do this, Machine Learning models were trained on the information in the ACN dataset, including information from the user-inputs. It was however, observed that the user-inputs could be inaccurate. Thus, another suitable application for Machine learning on the ACN dataset could be to attempt to predict the error in the user-inputs instead. Finally, as the continued electrification of the transport sector is set to continue and introduce a more widespread access to EV fast charging stations, it would be beneficial to analyse EV user-behaviour using data record at a fast charging station. With access to data recorded at the D.C. fast charging station available at the ACN, it could be interesting to analyse the differences in how the EV owners at the ACN utilise the fast charging station compared to the Level 2 charging stations. Similarly, a comparison of user-behaviour of fast charging stations and Level 2 charging stations could possibly be conducted with data from one of the other data sources that were presented in this thesis.

References

[1] Y. Amara-Ouali, Y. Goude, P. Massart, J.-M. Poggi and H. Yan. A Review of Electric Vehicle Load Open Data and Models. *Energies* 14 (8) (Apr. 2021). DOI: 10.3390/en14082233.

[2] E. Bitar and Y. Xu. Deadline Differentiated Pricing of Deferrable Electric Loads. *IEEE Transactions on Smart Grid* 8 (1) (2017). DOI: 10.1109/TSG.2016.2601914.

[3] A. L. Brenna. *Nytt rekordhøyt strømforbruk på morgenen (Retrieved 13.05.2021)*. Feb. 2021. URL: https://enerwe.no/stromforbruk/nytt-rekordhoyt-stromforbruk-pa-morgenen/395803.

[4] J. Burkin. *Irish Electric Vehicle Charge Point Status Datasets*. Oct. 2017. URL: http://www.mlopt.com/?p=6598.

[5] California Center for Jobs and the Economy. California Energy Price Data for December 2018 (Retrieved 24.05.2021). *California Center for Jobs and the Economy* (). URL: https://centerforjobs.org/ca/energy-reports/california-energy-price-data-for-december-2018.

[6] California Energy Commission. *California Energy Commission Zero Emission Vehicle and Infrastructure Statistics. (Retrieved 28.5.2021)*. 2021. URL: https://www.energy.ca.gov/data-reports/energy-insights/zero-emission-vehicle-and-charger-statistics.

[7] California Institute of Technology. *Caltech Academic Calendar (Retrieved 28.5.2021)*. 2021. URL: https://www.caltech.edu/academics/resources/academic-calendar.

[8] Ç. DERİCİOĞLU, E. YİRİK, E. ÜNAL, M. U. CUMA, B. ONUR and M. TÜMAY. A REVIEW OF CHARGING TECHNOLOGIES FOR COMMERCIAL ELECTRIC VEHICLES. *INTERNATIONAL JOURNAL OF ADVANCES ON AUTOMOTIVE AND TECHNOLOGY* (2018). DOI: 10.15659/ijaat.18.01.892.

[9] C. H. Dharmakeerthi, N. Mithulananthan and T. K. Saha. Impact of electric vehicle fast charging on power system voltage stability. *International Journal of Electrical Power and Energy Systems* 57 (2014). DOI: 10.1016/j.ijepes.2013.12.005.

[10] DNV GL and Poyry Management Consulting. Kostnader i stroemnettet - gevinster ved koordinert lading av elbiler. *The Norwegian Water Resources and Energy Directorate* (2019).

[11] ElaadNL. *ElaadNL shares new data sets demonstrating the rise of EVs and usage of Charging Stations (Retrieved 30.05.2021)*. Apr. 2020. URL: https://www.elaad.nl/news/elaadnl-shares-new-data-sets-demonstrating-the-rise-of-evs-and-usage-of-charging-stations/.

[12] esb.ie. ESB-E cars charge map (Retrieved 22.05.2021). *esb.ie* (2021). URL: https://esb.ie/ecars/charge-point-map.

[13] European Comission. *A European strategic long-term vision for a prosperous, modern, competitive and climate neutral economy (Retrieved 13.05.2021)*. Nov. 2018. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52018DC0773.

[14] M. C. Falvo, D. Sbordone, I. S. Bayram and M. Devetsikiotis. EV charging stations and modes: International standards. In: *2014 International Symposium on Power Electronics, Electrical Drives, Automation and Motion, SPEEDAM 2014*. 2014. DOI: 10.1109/SPEEDAM.2014.6872107.

[15] X. Hao, H. Wang, Z. Lin and M. Ouyang. Seasonal effects on electric vehicle energy consumption and driving range: A case study on personal, taxi, and ridesharing vehicles. *Journal of Cleaner Production* 249 (2020). DOI: 10.1016/j.jclepro.2019.119403.

[16] L. Held, A. Märtz, D. Krohn, J. Wirth, M. Zimmerlin, M. R. Suriyah, T. Leibfried, P. Jochem and W. Fichtner. The influence of electric vehicle charging on low voltage grids with characteristics typical for Germany. *World Electric Vehicle Journal* 10 (4) (2019). DOI: 10.3390/wevj10040088.

[17] J. D. Hunter. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* 9 (3) (2007). DOI: 10.1109/MCSE.2007.55.

[18] IONITY. *THE POWER OF 350 KW (Retrieved 21.05.2021)*. 2021. URL: https://ionity.eu/en/design-and-tech.html#.

[19] S. Johansson, J. Persson, S. Lazarou and A. Theocharis. Investigation of the impact of large-scale integration of electric vehicles for a Swedish distribution network. *Energies* 12 (24) (2019). DOI: 10.3390/en12244717.

[20] E. C. Kara, J. S. Macdonald, D. Black, M. Bérges, G. Hug and S. Kiliccote. Estimating the benefits of electric vehicle smart charging at non-residential locations: A data-driven approach. *Applied Energy* 155 (2015). DOI: 10.1016/j.apenergy.2015.05.072.

[21] H. H. Kvisle. The Norwegian charging station database for electromobility (NOBIL). *World Electric Vehicle Journal* 5 (3) (2012). DOI: 10.3390/wevj5030702.

[22] Z. J. Lee, T. Li and S. H. Low. ACN-Data: Analysis and applications of an open EV charging dataset. In: *e-Energy 2019 - Proceedings of the 10th ACM International Conference on Future Energy Systems*. 2019. DOI: 10.1145/3307772.3328313.

[23] Z. J. Lee, G. Lee, T. Lee, C. Jin, R. Lee, Z. Low, D. Chang, C. Ortega and S. H. Low. *Adaptive charging networks: A framework for smart electric vehicle charging*. 2020. DOI: 10.1109/TSG.2021.3074437.

[24] E. Lorentzen, P. Haugneland, C. Bu and E. Hauge. Charging infrastructure experiences in Norway -The worlds most advanced EV market. In: *EVS 2017 - 30th International Electric Vehicle Symposium and Exhibition*. 2017.

[25] A. Lucas, F. Bonavitacola, E. Kotsakis and G. Fulli. Grid harmonic impact of multiple electric vehicle fast charging. *Electric Power Systems Research* 127 (2015). DOI: 10.1016/j.epsr.2015.05.012.

[26] W. McKinney. Data Structures for Statistical Computing in Python. In: 2010. DOI: 10.25080/Majora-92bf1922-00a.

[27] Ministry of Climate and Environment. *Klimaendringer og norsk klimapolitikk (Retrieved 13.05.2021)*. 2020. URL: https://www.regjeringen.no/no/tema/klima-og-miljo/innsiktsartikler-klima-miljo/klimaendringer-og-norsk-klimapolitikk/id2636812/.

[28] Ministry of petroleum and energy. *Forskrift om leveringskvalitet i kraftsystemet (Retrieved 13.05.2021)*. Jan. 2007. URL: https://lovdata.no/dokument/SF/forskrift/2004-11-30-1557.

[29] Ministry of Transport and Communications. *Norge er elektrisk (Retrieved 13.05.2021)*. Nov. 2019. URL: https://www.regjeringen.no/no/tema/transport-og-kommunikasjon/veg_og_vegtrafikk/faktaartikler-vei-og-ts/norge-er-elektrisk/id2677481/.

[30] G. Newsom. *Executive Order N-33-20*. State of California, Mar. 2020. URL: https://www.gov.ca.gov/wp-content/uploads/2020/03/3.19.20-attested-EO-N-33-20-COVID-19-HEALTH-ORDER.pdf.

[31] F. S. Nilsen and S. O. Saeter. *Ny strømrekord i Norge (Retrieved 13.05.2021)*. Feb. 2021. URL: https://e24.no/olje-og-energi/i/6zBLm0/ny-stroemrekord-i-norge?referer=https%3A%2F%2Fwww.vg.no&fbclid=IwAR2jspL-opT8Q5mm2NmMoNupt05jF-JlflGqDhBUwA-9ig1KY1ptnFNUvmg.

[32] NOBIL. *Nobil charging stations, attribute types from database (Retrieved 13.05.2021)*. URL: https://nobil.no/admin/attributes.php.

[33] NOBIL. *NOBIL, the charging station database (Retrieved 09.05.2021)*. URL: https://info.nobil.no/eng.

[34] NOBIL.no. *NOBIL API Documentation (Retrieved 30.05.2021)*. Aug. 2018. URL: https://info.nobil.no/images/downloads/API-NOBIL-Documentation_v3_20180808.pdf.

[35] NVE. *Kraftnettet transporterer strømmen fra produsent til forbruker. Et velfungerende nett er en forutsetning for sikker strømforsyning. (Retrieved 21.05.2021)*. Oct. 2015. URL: https://www.nve.no/energiforsyning/nett/.

[36] F. Pedregosa FABIANPEDREGOSA, V. Michel, O. Grisel OLIVIERGRISEL, M. Blondel, P. Prettenhofer, R. Weiss, J. Vanderplas, D. Cournapeau, F. Pedregosa, G. Varoquaux, A. Gramfort, B. Thirion, O. Grisel, V. Dubourg, A. Passos, M. Brucher, M. Perrot andÉdouardand, a. Duchesnay and F. Duchesnay EDOUARDDUCHESNAY. *Scikit-learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, GRAMFORT ET AL. Matthieu Perrot*. Tech. rep. 2011.

[37] S. Raschka and V. Mirjalili. *Python Machine Learning: Machine Learning & Deep Learning with Python, Scikit-Learn and TensorFlow 2, Third Edition*. January 2010. 2019.

[38] T. F. Rosenbaum and D. A. Tirrell. *3/19/2020 - "Safer at Home" Order (Retrieved 23/05/2021)*. Mar. 2020. URL: https://together.caltech.edu/updates/general/3192020.

[39] S. Shahriar, A. R. Al-Ali, A. H. Osman, S. Dhou and M. Nijim. Machine learning approaches for EV charging behavior: A review. *IEEE Access* 8 (2020). DOI: 10.1109/ACCESS.2020.3023388.

[40] B. Shakerighadi, A. Anvari-Moghaddam, E. Ebrahimzadeh, F. Blaabjerg and C. L. Bak. A hierarchical game theoretical approach for energy management of electric vehicles and charging stations in smart grids. *IEEE Access* 6 (2018). DOI: 10.1109/ACCESS.2018.2878903.

[41] S. Srdic and S. Lukic. Toward Extreme Fast Charging: Challenges and Opportunities in Directly Connecting to Medium-Voltage Line. *IEEE Electrification Magazine* 7 (1) (2019). DOI: 10.1109/MELE.2018.2889547.

[42] SSB. *Transport står for 30 prosent av klimautslippene i Norge (Retrieved 13.05.2021)*. Aug. 2019. URL: https://www.ssb.no/natur-og-miljo/artikler-og-publikasjoner/transport-star-for-30-prosent-av-klimautslippene-i-norge.

[43] SSB. *Bilparken (Retrieved 13.05.2021)*. Mar. 2021. URL: https://www.ssb.no/bilreg.

[44] The pandas development team. *pandas-dev/pandas: Pandas*. Feb. 2020. URL: https://doi.org/10.5281/zenodo.3509134.

[45] UNFCCC. Paris Agreement. *Paris Agreement* (2015).

[46] G. van Rossum and F. L. Drake. *Python 3 Reference Manual*. 2009.

[47] A. von Meier. *Electric Power Systems: A Conceptual Introduction*. 2006. DOI: 10.1002/0470036427.

[48] M. Waskom. seaborn: statistical data visualization. *Journal of Open Source Software* 6 (60) (2021). DOI: 10.21105/joss.03021.

[49] T. Wildi. *Electrical Machines, Drives, and Power Systems*. Vol. 1. 2006.

[50] D. H. Wolpert. The Lack of a Priori Distinctions between Learning Algorithms. *Neural Computation* 8 (7) (1996). DOI: 10.1162/neco.1996.8.7.1341.

[51] T. Yuksel and J. J. Michalek. Effects of regional temperature on electric vehicle efficiency, range, and emissions in the united states. *Environmental Science and Technology* 49 (6) (2015). DOI: 10.1021/es505621s.

Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway