This paper was presented at The ISPIM Innovation Conference – Innovation, The Name of The Game, Stockholm, Sweden on 17-20 June 2018. The publication is available to ISPIM members at www.ispim.org.

# E-Mobility and Big Data – Data Utilization of Charging Operations

## Gerald Cuder *

Know-Center GmbH, Inffeldgasse 13, 8010 Graz, Austria.
E-mail: gcuder@know-center.at

## Gert Breitfuss, Roman Kern

Know-Center GmbH, Inffeldgasse 13, 8010 Graz, Austria.

Graz University of Technology, Institute of Interactive Systems and Data Science, Inffeldgasse 16c, 8010 Graz, Austria

## Roland Angerer

Has.To.Be GmbH, Salzburger Straße 20, 5550 Radstadt, Austria

* Corresponding author

**Abstract:** Electric vehicles have enjoyed a substantial growth in recent years. One essential part to ensure their success in the future is a well-developed and easy-to-use charging infrastructure. Charging stations themselves and any charging activity generates a lot of (big) data. The analysis for this data provides useful information about current E-Mobility behaviour. It bears the potential for informed decisions on the utilisation of the infrastructure and its future planning. In this paper we present data analytics methods and visualization technologies on such a data set collected by a large-scale, real-world charging infrastructure. To this end we researched established forecasting models like ARIMA and recent advances in machine learning for their usefulness in this setting. One objective of our research is, to provide a consumption forecast based on the historical consumption data. Based on this information, the operators of charging stations are able to optimize the energy supply. Furthermore, advanced prediction algorithms were applied to provide services to the end user regarding availability of charging stations. We managed to build models for most charge points solving all use cases with a comparably low error. The combined outcome of these analyses is of help for the infrastructure's operator to provide an improved service to its customers. The developed and tested solutions open up a broad agenda for future services based on charging operation data.

**Keywords:** Data-driven; electric vehicles; E-mobility; ARIMA; time series; travel planning; consumption prediction

## 1 Introduction

In recent years, electric vehicles (EVs) have become increasingly popular, as a rising number of car manufacturers will offer in future reasonably priced EVs for everyone. Many incentives, such as subsidisations, tax exemptions and free parking have been established that further stir the interest in EVs leading to an increased usage. Especially in European countries E-mobility has experienced a substantial grow since 2010 with both increased market shares and sales. In Norway the slope of the increase was progressed the most with a market share of 13% in the year 2014. By comparison, the market shares of following countries, Netherlands and Estonia, were with 1% quite far behind Norway (Figenbaum *et al.*, 2015). However, according to a recent survey (UBS, 2017), the market share of EVs will grow further until 2025 with a predicted share of 30% in Europe, whereas the most uprising economies of the world, China and USA, will be very far behind in 2025 according to forecasts (see Figure 1).

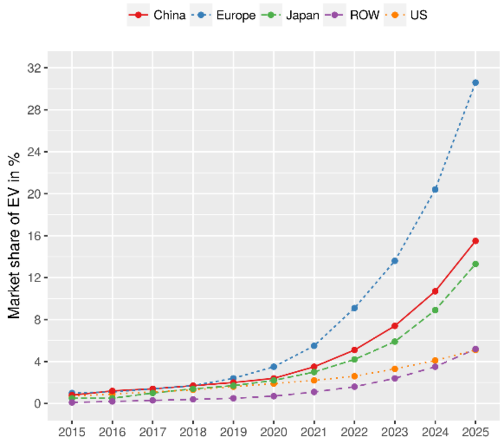


**Figure 1** Trend of market share for EV in different regions (UBS, 2017)

Another contributing factor for the observed increase in popularity of EVs is linked to the increase in charge points. The increase in density of the charge point network allows EV owners to travel further, since the most relevant perceived limitations of EVs is the range due to limited battery charge. According to (Figenbaum *et al.*, 2015), about 74% of the traditional car owners and at least 21% of the EV owners see one of the biggest disadvantages in the limited range of EV. Nowadays, provider of charging infrastructure (hardware and software) facilitate the set-up and operation of charge points (CP) for EV for single households and businesses. This will increase the number of CPs rapidly which might benefit the market share of EV as well, as a tighter and larger network of easy-to-use CPs increases the range of EVs drastically. An increasing share of EVs contributes to the local reduction of air pollutants produced by combustion engines especially in urban areas, which might be a lot more pronounced in the foreseeable future, according to a recent estimation for 2025 (Liberto *et al.*, 2017). Furthermore the overall soundscape of urban areas will undergo a substantial reduction due to nearly noiseless EV which might have a benefit for inhabitants of urban areas, (Genuit, 2013). As every charging at those CPs is recorded, a lot of data is generated and transferred to a central database. In this work we developed a framework based on the *Cross Industry Standard Process for Data Mining*

*(CRISP-DM)* (Wirth and Hipp, 2000) to analyse the data generated by charging stations and extract as many useful information in order to develop innovative services that benefits all involved parties equally. For instance, to allow the charge point operator (CPO) to implement a cost-optimized way to plan the energy supply for an individual CP, a sophisticated prediction strategy was developed to forecast a CP's both short- and long-term energy consumption. Moreover, the availability of a single charge point in terms of free connectors was also estimated based on historical charge events, as charging a EV usually takes much longer than to fuel a traditional car. So, the information on if a charge point is available or not might save the EV owner a lot of time, as he is able to plan his journey accordingly

Finally, the present paper illustrates, how the described use cases are solved with the use of state-of-the-art data analysis and machine learning technologies in order to extract data-driven features and services for charging station operators and end-users at the same time.

## 2 Related Work

Considering the fact that the boom of e-mobility is quite new and that the geographic density of CPs is steadily increasing, a comparably small amount of related work concerning data utilization of CPs and EVs has been conducted in the recent past. However, some work in a similar direction than ours has been done by Arias et al. (Arias and Bae, 2016) who used big data technologies for EV charging demand forecasting in South Korea by including real-world traffic data and weather data into their model. A similar approach was introduced in (Xydas et al., 2016) where the charging demand of EV in the UK was estimated by cluster and correlation analysis combined in a fuzzy based characterization model. However, both (Xydas et al., 2016) and (Arias and Bae, 2016) developed approaches that focused more on the energy supplier point of view, as the overall trend of the future energy demand caused by EV is estimated. In (Amini, Kargarian and Karabasoglu, 2016), they used an autoregressive integrated moving average (ARIMA) based approach to forecast both the charging demand of EVs and conventional electric load on the power grid by using historical load data and charging event data. Furthermore, an also very interesting approach was introduced by (Majidpour et al., 2015), who not only developed a way of generating very fast predictions of the overall consumption load of EV charging stations, but also addresses the sparsity of those time series. By transforming the time series problem into a supervised learning problem, they managed to create a forecast by applying fast machine learning (ML) algorithms such as k-nearest neighbours, on the one hand, and the historical average as a baseline, the other hand.

Concerning the optimization of travel time, (Qin and Zhang, 2011) have done some research by developing an algorithm that intelligently schedules car chargings by taking temporally and spatially charging activities into account.

## 3 Data description

Our dataset was provided by an Austrian independent e-mobility solution provider which offers a charging station management software solution This software solution allows private persons (households) and companies an easy set-up, operation and maintenance of one or more CPs. Although the data contain information on single charge events, the identity of the corresponding user remains secret, as the data was anonymized beforehand, the data generated by more than 1000 CPs between 2015 and 2017 have been analysed to extract as much useful information as possible. In total, we analysed 121,812 charging events of exactly 1,076 CPs. As stated in Table 1, the majority of these CPs are located in Germany and Austria, followed by Switzerland, Sweden and Spain. Moreover, the maximum timespan of charge recordings differed depending on the country, ranging from 765 days (Germany) to only 25 days (Slovakia).

**Table 1** Summary of charging station by country

| *Country* | *Number of CPs* | *Number of chargings* | *Total timespan (days)* |
|---|---|---|---|
| Germany | 719 | 86574 | 765.59 |
| Austria | 175 | 20621 | 757.42 |
| Switzerland | 100 | 7033 | 597.38 |
| Sweden | 52 | 4792 | 647.43 |
| Spain | 15 | 1833 | 582.34 |
| Croatia | 2 | 478 | 469.05 |
| Colombia | 1 | 185 | 214.41 |
| Slovakia | 1 | 113 | 24.99 |
| Italy | 3 | 76 | 462.16 |
| The Netherlands | 1 | 69 | 57.1 |
| Australia | 1 | 22 | 52.75 |
| China | 1 | 5 | 46.84 |
| Ireland | 2 | 4 | 22.09 |
| Monaco | 1 | 3 | 0.96 |
| Romania | 1 | 3 | 20.19 |
| Bosnia | 1 | 1 | 0.21 |
| **Total** | **1076** | **121812** | **43365** |

## 4 Methodology

In general, the data mining framework is a derivation of the *Cross Industry Standard Process for Data Mining (CRISP-DM)* (Wirth and Hipp, 2000) and mainly consists of following parts: (i) the *requirement analysis* and the *development of use cases*, (ii) the *data processing*, (iii) the *model approach* and (iv) the *visualization*. In (i) we conducted a structured requirement analysis with charge point operators in Austria, Germany and Switzerland in order to adapt and reshape predefined use case ideas. In (ii), the given data were prepared for the use cases by merging different data sources, cleaning and aggregation activities. In (iii) machine learning (ML) models were trained and evaluated in order to solve the defined use cases. Finally, the results were visualized in (iv) by integrating them into a specially developed web-based scalable visualisation tool. The whole workflow is shown in Figure *2*.

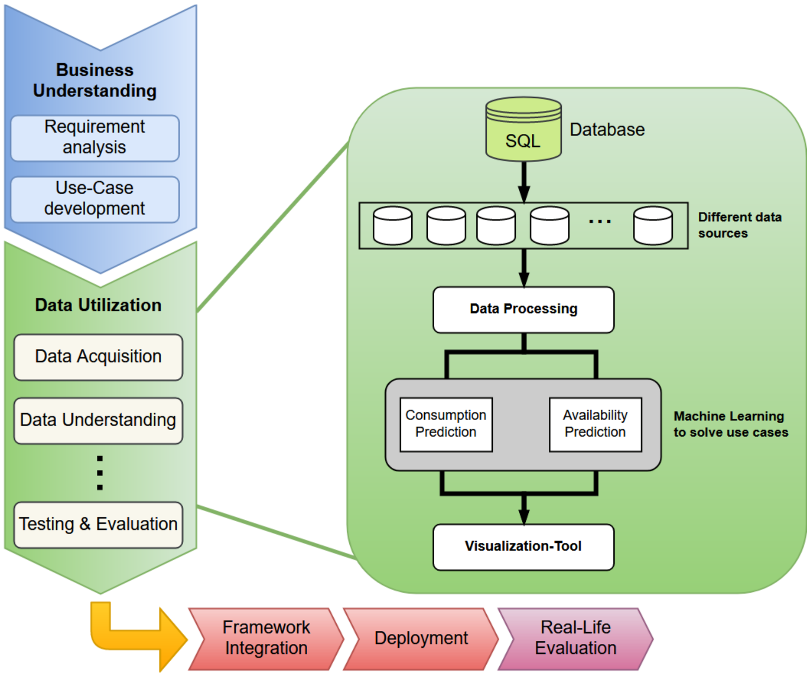


**Figure 2** Data utilization workflow describing the application of the CRISP-DM with the focus on *business understanding*, *data processing*, *machine learning for use cases* and *visualization* (Wirth and Hipp, 2000)

Further steps such as the *framework integration*, *deployment* and *real-life evaluation* were not in the scope of our research and are therefore not discussed in this paper.

### *Requirement Analysis*

In order to get a realistic business understanding a qualitative customer survey were carried out. Five personal interviews with charge point operators (2 in Austria, 2 in Germany, 1 in Switzerland) based on a semi-structured interviews guideline have been conducted. The focus of the qualitative survey was to obtain requirements and ideas on new data-driven reports and/or visualisations. The interviews were evaluated through qualitative content analysis by analysing the conversation logs systematically.

### *Use-Case Development*

After evaluating the outcome of the requirement analysis and the consideration of the inputs of data provider, we defined following use cases: *(i) Consumption Prediction* and a *(ii) Availability Prediction.* In *(i)* the general aim was to develop strategies for predicting the energy consumption of one or more CPs in a long- (next 7 days) and short-term (next 24 hours) scenario, while in *(ii)*, it was our goal to estimate the availability of a single charge point in an hourly manner. These use cases were chosen in order to gain the highest benefit to extend the existing operating system and to satisfy the needs of the CPO and the e-car owner.

### *Data processing*

Since the data generated by the charging stations contained a lot of large data sources, a sophisticated exploration and data processing strategy had to be developed which consisted of three consecutive phases, namely the combination phase followed by the cleaning and preparation phase. All data is stored in a MySQL database from which our framework directly acquires the required datasets. The data set itself spawns multiple data sources, therefore data fusion methods need to be applied. To this end, in the first step of the preprocessing pipeline, we aimed to combine as much information into consolidated datasets by selecting subsets from the individual data sources based on their unique information content. This way, we managed to reduce a wide range of different data sources down to only two datasets for further analysis, namely a dataset containing the energy consumption records of every charge event in a minutely interval, and another data set having information on every charge event (i.e. start time, end time). The resulting data sets had then to be preprocessed even further in the second step, the cleaning phase, by removing duplicate and erroneous entries and by imputing missing data points. Lastly, the datasets were prepared in order to be suitable to build further use cases on. First of all, the dataset containing the minutely time series of the charging events, was formatted into three differently shaped sets, by aggregating the energy consumption of each charge point over time, creating a minutely, an hourly and a daily time series of each charge point with the corresponding peak energy consumption in kilowatt hours (kWh). Second of all, we used sophisticated domain-specific feature engineering techniques in order to extract data-driven features for improving the model performance in later analysis. To all these steps, the inputs of the *requirement analysis* contributed specific domain knowledge in order to improve the whole process. Figure 3 gives a detailed overview on the pre-processing pipeline, while in Table 2 and Table 3 the two final datasets are described.

**Table 2** Information contained in the charge event dataset

| *Attribute name* | *Attribute description* |
|---|---|
| User | Unique ID for each user that connects to the CP |
| Charge point | Unique ID of the charge point |
| Start of charging | Timestamp (dd/mm/yyyy HH:MM:SS) when the user connects to the CP |
| End of charging | Timestamp (dd/mm/yyyy HH:MM:SS) when the user disconnects from the CP |
| Location | Longitude and latitude coordinates for the CP (WGS84) |

**Table 3** Information contained in the consumption dataset

| *Attribute name* | *Attribute description* |
|---|---|
| User | Unique ID for each user that connects to the CP |
| Charge point | Unique ID of the charge point |
| Timestamp | Timestamp (dd/mm/yyyy HH:MM:SS) when the energy consumption value is sent by a CP |
| Energy consumption value | Energy consumption value sent by the CP for a given timestamp |
| Location | Longitude and latitude coordinates for the CP (WGS84) |

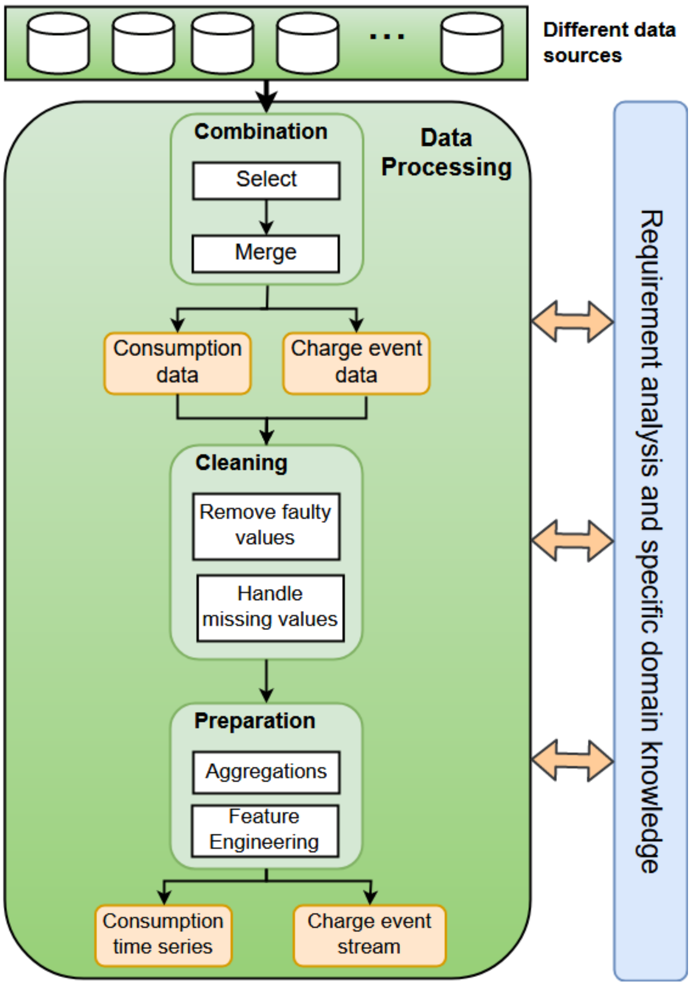


**Figure 3** Data processing workflow describing the way of the data from raw data to processed datasets ready for further analysis, with continuous input from the *requirement analysis* and specific domain knowledge

### *Consumption Prediction*

The objective concerning this use case was, to predict the energy consumption of a certain charge point for the next 24h and the next 7 days by using time series of past charging data. We dealt with this issue by treating it as a time series forecasting setting, consisting of the following steps. In the data pre-processing workflow (see Figure 3), the historical consumption data set, which contains consumption values in a minutely interval, is prepared by generating time series for each charge point with different aggregation levels. For the short-term prediction, time series in an hourly interval are created, while for the long-term forecast, a daily interval was considered sufficient. In order to solve this use case, we used separate optimized models for the long- and short-term forecast, as they could not be solved by using only one generalized model.

In general, for most of the charge points the data was very sparse with an average time of 22 hours between the charge events. This had to be overcome especially for the short-term prediction. Therefore, we used the historical average approach such as (Majidpour *et al.*, 2015) to generate a forecast for the next 24 hours. This means that the forecast of the energy consumption value for a given charging station on for a specific hour depends on the mean of all past energy consumption values on hourly values every 24 hours. Using this approach, we ensured that even the lack of stationarity of time series did not affect the forecast, on the one hand, and that this also could be applied to newly built charging stations with a relatively small amount of data available with increasing forecast accuracy as the CP is used more often. In order to validate the predictive performance of the algorithm, we made a train-test-split, using the last three weeks for validation. The model was built on the rest of the data set (train) making sure not to include and information from the test set in the training phase. Following equation shows how a historical average model is built, where $D$ stands for the number of historical data points used for the estimation.

$$\hat{y}(t) = \frac{1}{D}\sum_{d=1}^{D} y(t-24d)$$

Concerning the long-term prediction for the next 7 days in a daily interval, several established time series models have been applied, including autoregressive integrated moving average models (ARIMA), exponential smoothing (ETS) and a linear model that decomposes the time series into trend and season. Especially ARIMA and ETS models have often been used in forecasting of energy consumption, mostly, however, of whole cities and households (Al-Musaylh *et al.*, 2018; de Oliveira and Cyrino Oliveira, 2018). ARIMA combines auto-regressive (AR) and moving average model approaches with additional differencing in order to establish stationarity and predicts future values by learning from the past (Chatfield, 2016). ARIMA models are usually used to capture seasonalities (recurring values within a certain period) in the data, which is mostly the case in series of energy consumption values, as load patterns are highly dependent on the time of the day (Taylor, 2003).

### *Availability Prediction*

Approximately 62% of traditional car owners and 22% of EV owners see the accessibility of CP as their biggest drawback. Motivated by this, we considered to develop a use case that aims to tackle this disadvantage by calculating an estimation for each charge point giving the EV owners the opportunity to choose the best time for charging their cars. This was achieved by using the historic average approach, applied on the charge event dataset which stores information on every charging, as stated in Table 2 Information contained in the charge event data The prediction of a CP's availability is calculated hourly, resulting in estimation for every hour of the day. Additionally, two different weighting strategies that individually value each charge event have been implemented. This has been done due to the structure of the data, having a varying number of charging events per month over the overall time range. Two different patterns were identified: Many CPs have fewer charging events in the past compared to newer data, while some CPs have a varying density of charge events, regardless of the time in the data. As both issues can seriously bias the final estimation of the CP's availability, a *(i) time-based* and a *(ii) density-based weighting strategy* were introduced, reducing the bias in the overall estimation for each CP individually.

In *(i)*, a two-stage weighing approach is applied which first weights data from each year separately, whereas newer years are valued the most. In the second step, the last three months of the data are weighted additionally. These weights were estimated heuristically based on domain expertise. This approach addresses the issue, that newer data is much more valuable than older data, by setting the weights for older charge events lower compared to newer ones. This strategy follows the principle of *exponential smoothing* (Box *et al.*, 2015), however we applied it in a two-step approach. The approach in *(ii)* follows a different strategy: Instead of addressing the topicality of the charge event, it values individual charging events in months with a higher charging density, regardless of the year the charging took place. The latter approach has a major advantage compared to the former one, as it also can be applied to CPs which has only a few months of data available.

### *Visualization framework and applied technologies*

In order to create an interactive and scalable platform to visualize the results of all use cases, the R-shiny framework (Chang *et al.*, 2017) was used. Using an interactive framework such as *shiny* allowed us to create highly sophisticated visualizations that are scalable and zoomable. This was necessary, as we had to deal with data of nearly two years, on the one hand, and there were over a thousand CPs to visualize on the other hand. Within the application, the user is able to interactively select the desired CP for each particular use case which makes the operation of multiple CPs a lot less complicated for the owner.

Besides the *shiny* framework, we used *R, a programming language for statistical computing* (R Core Team, 2017), to solve the use cases and to do the data processing. Additionally, we chose MySQL for data storage, as it comes with a lot of useful tools to handle many data sources.

## 5 Evaluation

In order evaluate our results, we did not use the whole datasets to build our models. Instead, we applied classic validation strategies used in state-of-the-art modelling of timeseries. ARIMA and all other time series models were validated using the rolling forecast method (Hyndman and Athanasopoulos, 2018) which is similar to k-fold cross validation in traditional machine learning. The algorithm starts with an initial train set which size is defined by the user. The corresponding test set is the one observation next to the last one in the train set. The model is trained and then tested on the single test observation. Then, the test observation is added to the train set and the model is retrained and again tested on the next observation which is not in the train set. This process is iterated till no observations are left to test the model on. Figure 4 summarizes the principle of the rolling forecast validation, whereas train and test set are represented by the blue and red points, respectively.

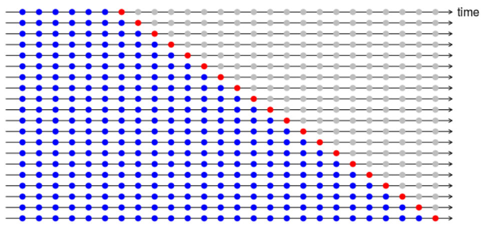


**Figure 4** Rolling forecast method as stated in (Hyndman and Athanasopoulos, 2018)

The error of both approaches was estimated by calculating the root mean squared error (RMSE), whereas $\hat{y}_t$ is the predicted value and $y_t$ the actual value at time $t$. Thus the RMSE describes the standard deviation between the predicted and the actual value over several predictions $n$ (Chatfield, 2016; Hyndman and Athanasopoulos, 2018).

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(\hat{y}_t - y_t)^2}{n}}$$

For the evaluation of the availability prediction, we used a similar approach, but instead of using only the next day, we tested the model on the whole next week in the validation set. The performance of the model was then estimated by calculating the mean percentage error (MPE) on the test set for each day with and without applied weighting strategies, respectively. The error was calculated like following:

$$MPE = \frac{100\%}{n}\sum_{t=1}^{n}\frac{y_t - \hat{y}_t}{y_t}$$

## 6 Results

This section describes the most relevant results gathered from the use cases. Figure 5 and Figure 6 show the results of UC 1 which deals with the forecast of a CP's energy consumption, while Figure 7 shows the results of the availability prediction. For demonstrative purpose, the following figures all represent a single CP (located in Graz, Austria).

As it is shown in Figure 5, the daily prediction for the next 7 days using an ARIMA model worked quite well producing a valid forecast (left-hand side) with a very low corresponding validation error (right-hand side). In the graphic, the red line indicates the actual energy consumption while the blue line stands for the validation error. The green line and the shaded areas shows the 7-day prediction with the corresponding prediction intervals.

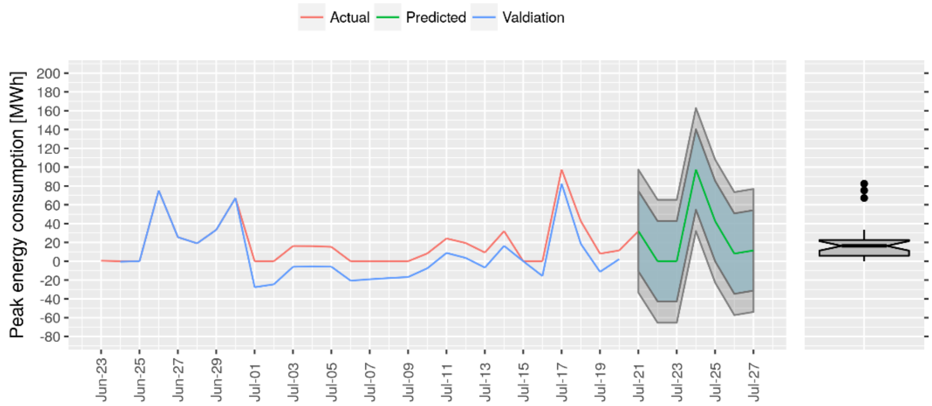


**Figure 5** Daily consumption prediction using ARIMA

While the previous figure gave an example for the long-term forecast, Figure 6 shows how the short-term forecast performs using the historical average on hourly data. On the upper left-hand side of the figure, the forecast for the next 24 hours is shown, while on the right-hand side the validation error on the test dataset indicates that the forecast to be quite accurate. The lower part of the figure visualizes how the historical average model performed on the test dataset by comparing the actual data with the model output.

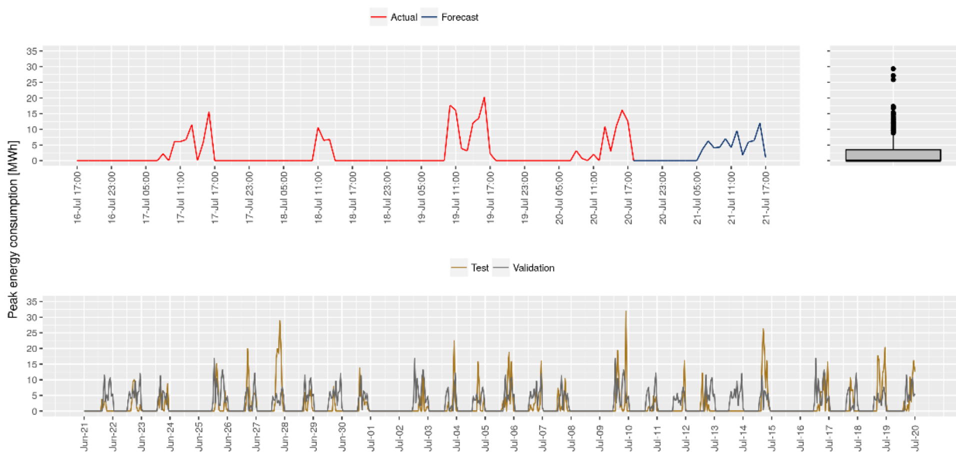


**Figure 6** Hourly consumption forecast using the historical average

Figure 7 describes the results of the third use case where an estimation of a CP's availability has been developed using again the historic average on charge event data. The figure shows the estimated load in percent over all hours of the day in combination with the estimated average time how long people usually spend there. On the right-hand side of the figure, the validation error was computed showing an error in the range from 9% to 13% depending on the weighting strategy applied. It also shows that the density-based weighting strategy works best having the lowest validation error.

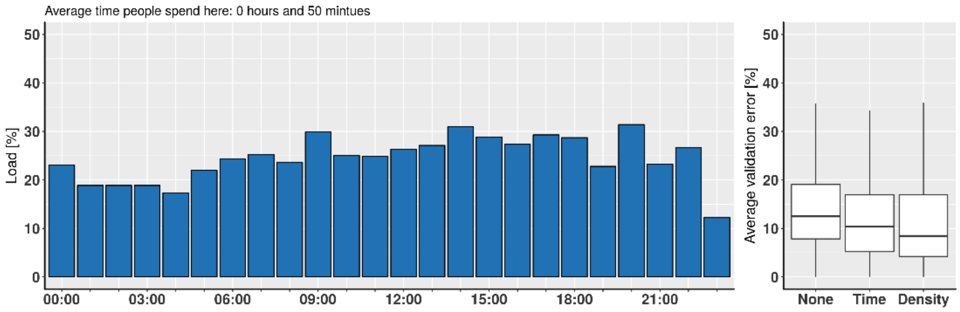


**Figure 7** Estimation of a CP's availability using the historical average of charge events

## 7 Discussion

E-mobility is an upcoming field and is going to affect the lives of all of us in the foreseeable future. With the advancement of EVs and charging stations, more and more data is produced and can be utilized for research and service development. In this work, we have developed a workflow that both cleans and facilitates data generated by more than thousand CPs in order to produce outputs that are useful to all parties involved. Our research not only considers an isolated subset of CPs within a high spatial proximity, (Arias and Bae, 2016; Xydas et al., 2016), using several external data sources like traffic and weather data. Hence, the mentioned approaches are more focused on the future influence of e-mobility on the overall power grid rather than predicting the power consumption of each individual CP within their data. We assume that the developed approaches can be applied to any CP regardless of its location. Our approach gains an immense benefit for all parties involved. First of all, the existing version of the operating system can now be extended by implementing an intelligent data pre-processing pipeline providing a powerful backend to provide useful features, such as a short- and long-term prediction of a CPs power consumption

Other parties involved in the environment of e-mobility also benefit from the outcome of this work. For instance, the consumption prediction allows the operator of a charge point to establish a cost-optimized energy supply, as violation of the contractually guaranteed peak power results in large monetary fines. On the contrary, knowing the future energy demand of a CP allows the charge point operator to adapt the contract with the local energy supplier at short notice in order to avoid those fines. From the e-car owners point of view, the availability prediction seems to be most beneficial, as charging an EV still consumes a considerable amount of time, depending on the charging technology and the e-car itself, respectively. Knowing the time, when a CP is most likely to be available, helps to reduce at least unnecessary waiting times and enables e-car users to optimize their travel strategy. Furthermore, optimizing travel planning might result in a higher probability that people who still use a traditional car start considering buying an EV, as the majority of non EV owners still see in the access to CPs (62%) on the one hand, and in the time it takes an EV to charge (51%), on the other hand the biggest disadvantage in e-mobility (Figenbaum *et al.*, 2015). The developed weighing strategies improve the availability prediction even further, however further validation will be necessary.

## 8 Conclusion and outlook

In this work, we proposed a framework for data utilization in order to extract data-driven features from CP data streams. The technological challenges could be solved by advanced data processing, by combining several data sources and by development of sophisticated strategies to weight historical data. The presented work shows how (big) data analytics and data-driven business can lead to service-innovation in the e-mobility domain. By using a two-way approach (customer and data focus) two use cases were identified for further development: to provide consumption forecast and to determine prediction information on charge point availability for EV owners. Providing a sophisticated solution for every of these use cases, we ensured that the charge point operator and the EV owner gained as much benefit as possible from the existing data.

Besides the benefit that has already been gained from the current data, we are also going to do further research, as the partnership with our partner is aimed to be long-term. As CP data is highly time dependent, we get updates of our database on a regular basis which extends the scope of our research sequentially. From our side, there are lots of suggestions of how the data utilization framework proposed in this work could be extended. For instance, if more data become available, the consumption can be predicted using Deep Learning models such as Long-Short-Term-Memory (LSTM) recurrent neural networks which have been successfully applied to time series forecasting recently (Kong *et al.*, 2018). Furthermore, we plan to develop a spatial regression model that estimates the optimal location to build a new CP by gathering several external data sources, such as traffic data, weather data and demographic data. Another hot topic is located in the security domain, as RFID cards used to identify a user at a certain CP can easily be copied and are therefore prone to fraud. We plan to establish a fraud detection system similar to those already used in the financial domain (credit cards), however with domain specific adoptions to e-mobility. Furthermore predictive maintenance is a major goal in every field of industry 4.0, as the reduction of downtimes increases the productivity and decreases unnecessary high costs for maintenance. Depending on the industry, maintenance can make out up to 60% of the production costs (Mobley, 2002; Lu, Durocher and Stemper, 2009). Thus we are eager to apply a certain predictive maintenance strategy for CPs, as the reduction of downtimes bear a huge benefit for the CPO.

## References and Notes

Al-Musaylh, M. S. *et al.* (2018) 'Short-term electricity demand forecasting with MARS, SVR and ARIMA models using aggregated demand data in Queensland, Australia', *Advanced Engineering Informatics*. Elsevier, 35, pp. 1–16. doi: 10.1016/J.AEI.2017.11.002.

Amini, M. H., Kargarian, A. and Karabasoglu, O. (2016) 'ARIMA-based decoupled time series forecasting of electric vehicle charging demand for stochastic power system operation', *Electric Power Systems Research*. Elsevier B.V., 140, pp. 378–390. doi: 10.1016/j.epsr.2016.06.003.

Arias, M. B. and Bae, S. (2016) 'Electric vehicle charging demand forecasting model based on big data technologies', *Applied Energy*. Elsevier Ltd, 183, pp. 327–339. doi: 10.1016/j.apenergy.2016.08.080.

Box, G. E. P. *et al.* (2015) *Time series analysis : forecasting and control.*

Chang, W. *et al.* (2017) 'shiny: Web Application Framework for R'. Available at: https://cran.r-project.org/package=shiny.

Chatfield, C. (2016) *The Analysis of Time Series: An Introduction, Sixth Edition*. CRC Press (Chapman & Hall/CRC Texts in Statistical Science).

Figenbaum, E. *et al.* (2015) 'Increasing the competitiveness of e-vehicles in Europe', *European Transport Research Review*, 7(3), p. 28. doi: 10.1007/s12544-015-0177-1.

Genuit, K. (2013) 'What will be the influence of e-mobility on soundscape?', in *Proceedings of Meetings on Acoustics*. Acoustical Society of America, pp. 040053–040053. doi: 10.1121/1.4800426.

Hyndman, R. and Athanasopoulos, G. (2018) *Forecasting: principles and practice*. 2nd edn. OTexts.

Kong, W. *et al.* (2018) 'Short-Term Residential Load Forecasting Based on Resident Behaviour Learning', *IEEE Transactions on Power Systems*, 33(1), pp. 1087–1088. doi: 10.1109/TPWRS.2017.2688178.

Liberto, C. *et al.* (2017) 'Evaluation of the impact of e-mobility scenarios in large urban areas', in *2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*. IEEE, pp. 379–384. doi: 10.1109/MTITS.2017.8005701.

Lu, B., Durocher, D. B. and Stemper, P. (2009) 'Predictive maintenance techniques', *IEEE Industry Applications Magazine*, 15(6), pp. 52–60. doi: 10.1109/MIAS.2009.934444.

Majidpour, M. *et al.* (2015) 'Fast prediction for sparse time series: Demand forecast of EV charging stations for cell phone applications', *IEEE Transactions on Industrial Informatics*, 11(1), pp. 242–250. doi: 10.1109/TII.2014.2374993.

Mobley, R. K. (2002) *An introduction to predictive maintenance*. Butterworth-Heinemann.

de Oliveira, E. M. and Cyrino Oliveira, F. L. (2018) 'Forecasting mid-long term electric energy consumption through bagging ARIMA and exponential smoothing methods', *Energy*. Pergamon, 144, pp. 776–788. doi: 10.1016/J.ENERGY.2017.12.049.

Qin, H. and Zhang, W. (2011) 'Charging scheduling with minimal waiting in a network of electric vehicles and charging stations', *Proceedings of the Eighth ACM international workshop on Vehicular inter-networking - VANET '11*, p. 51. doi: 10.1145/2030698.2030706.

R Core Team (2017) 'R: A Language and Environment for Statistical Computing'. Vienna, Austria. Available at: https://www.r-project.org/.

Taylor, J. W. (2003) 'Short-term electricity demand forecasting using double seasonal exponential smoothing', *Journal of the Operational Research Society*. Palgrave Macmillan UK, 54(8), pp. 799–805. doi: 10.1057/palgrave.jors.2601589.

UBS (2017) 'Q-Series UBS Evidence Lab Electric Car Teardown – Disruption Ahead?' Available at: http://www.advantagelithium.com/_resources/pdf/UBS-Article.pdf (Accessed: 13 March 2018).

Wirth, R. and Hipp, J. (2000) 'CRISP-DM: Towards a standard process model for data mining', *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, pp. 29--39. Available at: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.198.5133 (Accessed: 23 April 2018).

Xydas, E. *et al.* (2016) 'A data-driven approach for characterising the charging demand of electric vehicles: A UK case study', *Applied Energy*. Elsevier Ltd, 162, pp. 763–771. doi: 10.1016/j.apenergy.2015.10.151.