# A Computational Framework for Managing Electric Vehicle Charging Infrastructure

[1*]Dario Pevec, [2]Martin A. Kayser, [3]Jurica Babić, [4]Arthur Carvalho, [5]Yashar Ghiassi-Farrokhfal, [6]Vedran Podobnik

[1,3,6] University of Zagreb, Faculty of Electrical Engineering and Computing, Croatia
[2,5] Erasmus University, Rotterdam School of Management, Netherlands
[4] Miami University, Farmer School of Business, USA

* E-mail: dario.pevec@fer.hr

**Abstract**

Current trends suggest that there is an increase in the overall usage of electric vehicles (EV). This, in turn, is causing drastic changes in the transportation industry and, more broadly, in business, policymaking and society. One concrete challenge brought by the increase in the number of EVs is a higher demand for charging stations. This paper presents a computational framework that uses real-world data to answer questions related to EV charging infrastructure, such as where to place new chargers and how many chargers are needed to bring energy utilization to a desirable level. Our framework allows one to predict charging station utilization even when EV charging infrastructure and/or contextual data change. We foresee that the proposed framework can be used by EV charging infrastructure providers as a decision support tool that prescribes an optimal area to place a new charging station.

**Keywords:** Electric vehicles, green transportation, charging infrastructure, big data analytics, computational framework

## I. Introduction and Related Work

$CO_2$ emissions are considered one of the prime factors behind climate change. The transportation sector, in particular, is one of the main contributors to $CO_2$ emissions. This clearly implies that a possible solution to reduce such emissions is to invest in *green transportation*, such as electric vehicles (Saber and Venayagamoorthy, 2011). EV penetration is generally growing in developed markets, *e.g.,* the EV penetration in the Netherlands has increased nine-fold in the four-year period between 2013 and 2016 (National Enterprising Netherlands, 2017).

Behind such a growth is a better understanding of the factors that influence (potential) EV owners' preferences and behavior. For example, Adnan et al. (2016) developed a theoretical framework for adoption behaviour with the ultimate goal of understanding the driving factors related to EV adoption. An important factor that negatively impacts (potential) EV owners is now called *range anxiety*, which is defined as the fear of running out of electricity before reaching an available (*i.e.*, unoccupied) charging station (CS) (*e.g.,* see the work by Neubauer and Wood, 2014).

The above setting leads us to formulate the following research question: *"Where should an EV charging infrastructure provider place a new CS?"* Answering this question is a very important step towards a higher acceptance of EVs since an increase in the number of charging stations, as well as their strategic geographic distribution, can significantly lower an EV owner's range anxiety and, consequently, have a positive impact on EV adoption.

In this paper, we provide an answer to our research question using predictive models trained on real-world data. Our ultimate goal is to build a generic computational framework capable of predicting where a charging-infrastructure provider should place a new charger to efficiently meet the current demand for charging stations. This shall be done by taking into account the whole network of EV charging stations in a certain region. Besides the predictive aspect, the proposed framework should be able to report key insights on EV owners' charging behavior based on the underlying dataset provided as input.

While research papers on this topic already exist, they substantially differ from the approach we are proposing in our work. For example, Bikcora et al. (2016) suggested an approach to forecast the availability of charging stations and their charging rates. Although these authors used similar data as we do, their ultimate goals were not the same as ours.

Develder et al. (2016) analysed the charging behaviour of EV owners based on data from 2009 to 2015. Their approach was to cluster charging sessions based on arrival and departure times. The obtained clusters were later labelled as *park to charge*, *charge near work*, or *charge near home*. We note, however, that such a clustering is not our ultimate goal.

Yi and Bauer (2016) proposed a framework for optimal placement of charging stations by using heuristic approaches. They tested their framework in simulations as well as in a real-life scenario (*i.e.*, the city of Chicago). The suggested algorithms to maximize the number of reachable households under certain constraints and to minimize energy transportation reached satisfactory performance for strategic placement of charging stations. There are other research projects that study the impact on the energy grid caused by the increase in the number of EV charging stations, *e.g.,* see the work by Qian et al. 2010 and Aabrandt et al. 2012.

An interesting aspect of our work is that we base our results purely on real-life data. Some researchers, alternatively, use simulated data. Xi et al. (2013), for example, modelled a charging infrastructure based on simulation results from assumed flows of EVs. Babic et al. (2015) developed a simulation model for smart parking lot. Their parking lot data in their simulations are derived from real-life data, but the behavior of EV owners is simulated. Valogianni et. al. (2015) dealt with the problem of coordination mechanisms for variable-rate EV charging using agent-based simulations.

Another important research stream in this field is about the interactions between EVs and the energy grid. Kahlen and Ketter (2015) developed an algorithm that can determine when a fleet of EVs should absorb/provide energy from/to the grid. Ketter et. al. (2016) explained how competitive benchmarking can be a valuable research tool in sustainable energy.

This paper has four other sections besides this introductory section. Section 2 describes our framework and underlying data. Section 3 describes our main findings. We discuss about our results and the applicability of our framework in Section 4. Finally, we conclude in Section 5.

## II. Approach: Real-world Data-driven Machine Learning

To answer our research research question, we develop a framework used to predict the charging utilization in a certain region. That information can then be used to help an infrastructure provider when deciding about the deployment of new charging stations. Our framework relies on very specific data sets. We next describe the process of assembling such a dataset (depicted in Figure 1) before providing further details on our *Computational Framework for Extending Electric Vehicle Charging Infrastructure* (EVCI Framework).

### *II.1. Dataset Pre-processing*

The main data set we use in our analysis was provided by ElaadNL, one of the major charging infrastructure providers in the Netherlands. The dataset provided by ElaadNL contains all the charging transactions from the beginning of 2013 until the end of 2016 for their charging infrastructure. Some of the variables in the data set are *Transaction ID*, *Charge Point ID*, *Transaction start/end Time, Connected time, Charge time,* and *Idle time*.

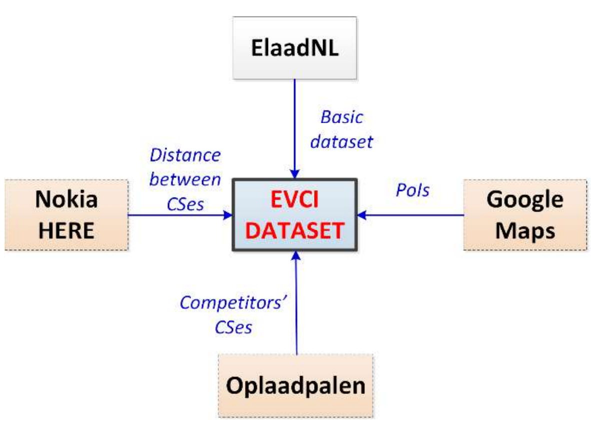



Fig. 1: EVCI Dataset

Tab. 1: Number of transactions and charging plugs in our data set per year

| Year | Number of transactions | Number of chargers |
|---|---|---|
| **2013** | 165,641 | 2,676 |
| **2014** | 342,419 | 2,687 |
| **2015** | 391,375 | 2,728 |
| **2016** | 548,317 | 2,922 |
| **CAGR** | 34.89 % | 2.22 % |

A transaction is defined by all the actions from the time when an EV owner plugs the EV to the charger until the EV is unplugged. The number of charging points maintained by ElaadNL and the number of transactions per year are shown in Table 1. It should come as no surprise that the number of transactions is growing at a much faster rate than the number of charging points, as one can see from the calculated CAGR (*Compound Annual Growth Rate*) values.

After processing the dataset into a format suitable for further analysis, we end up with the following variables:

- Start and end times of each charging session (i.e. month, day, hour, minute);
- EV parking duration (in minutes);
- EV charging duration (in minutes);
- Charging station cluster identifier;
- Number of chargers for the charging cluster when the charging took place.

As we elaborate later on, charging stations were grouped into clusters based on their geo-location. The initial ElaadNL dataset was extended with information about *places of interest (PoIs)* distributed across 13 categories (see Table 2). For a given charging event, we considered PoIs in a 500-meter radius to the charging station since PoIs located within that distance have significant influence on the charger utilization (Wagner et al. 2013.). Information about PoIs was gathered using the *Google Maps API*.

Tab. 2: PoI categories and descriptions

| PoI Category Number | PoI Category |
|---|---|
| Cat 1 | Community |
| Cat 2 | Drinks |
| Cat 3 | Entertainment |
| Cat 4 | Finance |
| Cat 5 | Food |
| Cat 6 | Health |
| Cat 7 | Office |
| Cat 8 | Religion |
| Cat 9 | School |
| Cat 10 | Shop |
| Cat 11 | Shop Essentials |
| Cat 12 | Sport |
| Cat 13 | Transport |

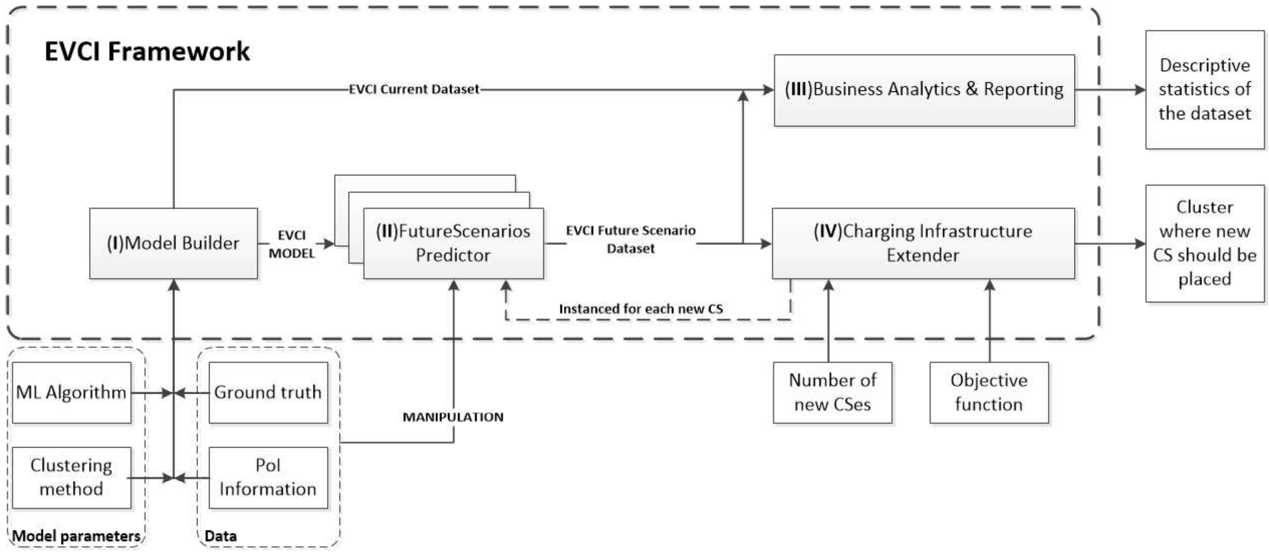



Fig. 2: EVCI Framework composition

Since ElaadNL is not the only provider of charging infrastructure in Netherlands (it has about 15% of the total infrastructure), the number of competitors' charging stations was also added to the dataset for each charging-station cluster. The reason behind this lies in the fact that such information is highly important for prediction of charger utilization since the more alternatives a user has, the higher the probability he/she will choose one of the competitors' charging station. Information about all charging stations in the Netherlands is available through the *Oplaadpalen API* (Oplaadpalen 2017). In our dataset, competitors' charging stations are treated as another category of PoIs.

### *II.2. EVCI Framework*

The main contribution of the presented research is the "*Computational Framework for Extending Electric Vehicle Charging Infrastructure*" (EVCI Framework). The EVCI Framework is a computational artefact that can be used to predict the utilization of an EV charging infrastructure. Such an information can be used, for example, to identify a charging station that would benefit most from an added charger. The framework itself is generic and adaptive with various parameters that can be manipulated, *e.g.,* the charging-station clustering method, the clustering distance threshold, the machine-learning method, *etc*. The framework composition is depicted in Figure 2. We now describe the framework's four modules: *EVCI Model Builder, EVCI Future Scenarios Predictor, EVCI Business Analytics & Reporting,* and *EVCI Charging Infrastructure Extender module.*

**EVCI Model Builder module** is the part of the EVCI Framework that *transforms* the dataset into a more appropriate format for analysis. For example, it clusters the charging stations and calculates the utilization for each cluster. Furthermore, this module creates and fine-tunes the prediction model. The inputs to this module are the *ground truth dataset* and *PoI information*, as well as information about the chosen *clustering method* and *machine learning (ML) algorithm*. The outputs from this module are: (i) *EVCI Current Dataset* (*i.e.,* dataset with calculated utilization) which can further be used for analysis and reporting of the ground truth scenario); and (ii) *EVCI Model*, *i.e.*, a machine learning model that can further be used for sensitivity analysis of various future scenarios.

**EVCI Future Scenarios Predictor module** performs sensitivity analysis of various future scenarios based on the EVCI model. This module enables manipulation with input data (*e.g.,* changing the number of available CSes in a certain CS cluster, adding new PoIs in a vicinity of certain CSes, changing EV penetration) and prediction of utilization based on that manipulation. Inputs in this module are manipulated *ground truth dataset* and *PoI information*. Output of this module is the EVCI Future Scenario Dataset (*i.e.*, dataset with predicted utilization values) which can further be used for analysis and reporting of the future scenarios.

**EVCI Business Analytics & Reporting module.** Module for statistical analysis of the dataset and comparison of current state with the predicted state of the utilization. Input in this module are *ground truth* and *dataset with predicted values*. Based on the provided data, module outputs *descriptive statistics of the dataset*. Descriptive statistics can be applied to different levels of granularity for each year (*e.g.,* month, year, day, *etc.*). Additionally, longitudinal comparison of data through years can be performed (Subsection ***III.2.***).

**EVCI Charging Infrastructure Extender module** uses algorithm for deciding optimal placement of new CSes and calculates difference between utilization of CSes in ground truth and new dataset where number of CSes is increased in each cluster. Based on that difference and the objective function, algorithm decides which cluster would benefit the most from deploying new CSes and outputs it as the result. This module instances new *Future Scenarios Predictor module* for each new available CS and in each iteration, decides where CS should be deployed.

## III. Main findings

This section describes and explains four main findings observed in the process of development of the EVCI framework as well as in the results of the EVCI Framework calculations.

### *III.1. Clustering methods*

Quality of the proposed prediction framework is not only dependent on the quality of data and prediction algorithm employed but it is also highly dependent on the clustering method used for grouping CSes. This finding is related to the EVCI Model Builder module, since the ML model it produces is directly dependent on the quality of the data, algorithm employed and clusters created. Our framework uses a distance-based method for the CS clustering, but instead of the traditional approach which measures aerial distance our approach uses the *Nokia HERE API* to take into consideration the driving distance between each pair of CSes. This approach improved the distance accuracy by 31% (in comparison to the traditional aerial distance) for our dataset. Furthermore, while the aerial distance grouping can result in unrealistic clusters (*e.g.,* grouping together CSes around the bay that have small aerial distances, but are not directly connected via road), usage of driving distance approach prevents such scenarios. The comparison of two described clustering approaches is depicted in Figure 3 – the approach based on driving distance results in more clusters, what is expected since aerial distance is shorter than driving distance.

Our clusters are based on 3 km distance because we assumed that an average owner of an EV has similar behaviour as an owner of a fossil-fuelled vehicle and is willing to travel between 5 and 10 minutes to reach next available CS. Taking furthermore into account a fact that an average speed in a populated area is between 20 and 35 km/h (Langer and McRae 2013, Statista 2017), we can assume that the EV owner is willing to travel approximately 3 km to reach the next available CS. This method could be problematic as it could result in longitudinal clusters along highways. Solution for that problem is a hybrid approach with distance based clustering and end-to-end cluster radius limit which would prevent clusters that spans more than the defined radius.

### *III.2. Descriptive statistics of the dataset*

Another main finding of our framework is based on the descriptive statistics of the dataset (output of the *Business Analytics & Reporting module*) which leads us to conclusions related to *utilization of CSes and parking spots* described in the following paragraphs (please note that utilization of CSes and utilization of parking spots are two different key performance indicators: *utilization of a CS* is defined as the percentage of time spent on charging, while *utilization of a parking spot* is defined as the percentage of a total time parking spot with charger was occupied).

As expected, from year to year utilizations of both parking spaces and charging plugs are increasing (Table 3). Figure 4 depicts daily charging and parking service utilizations through the year for all years available (*i.e.,* 2013-2016).

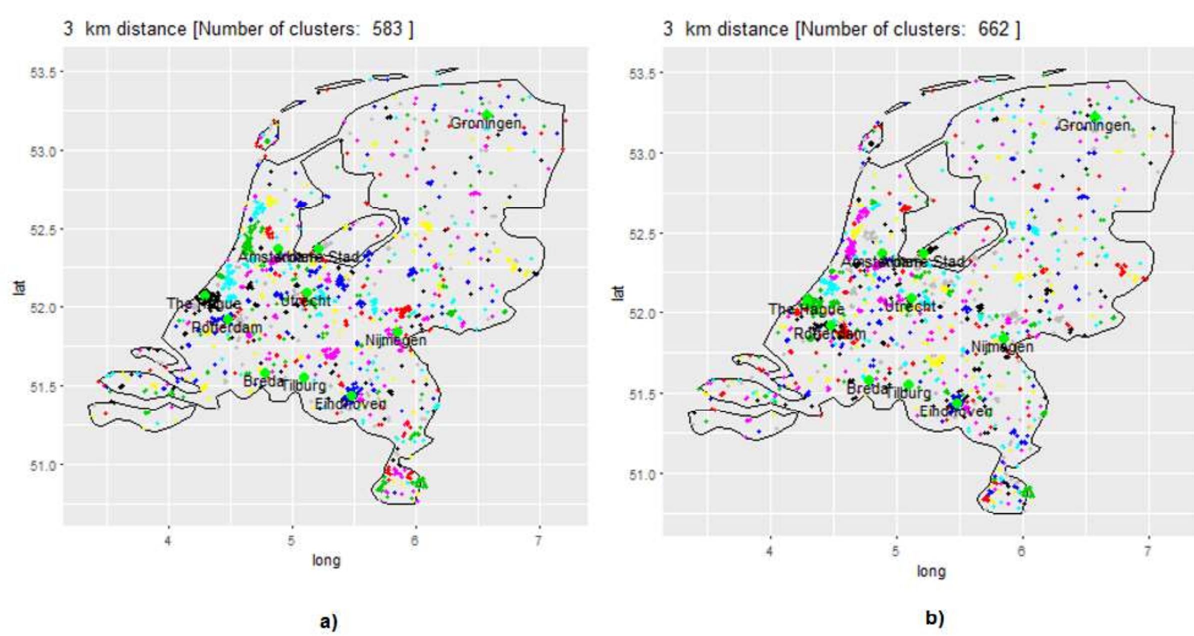
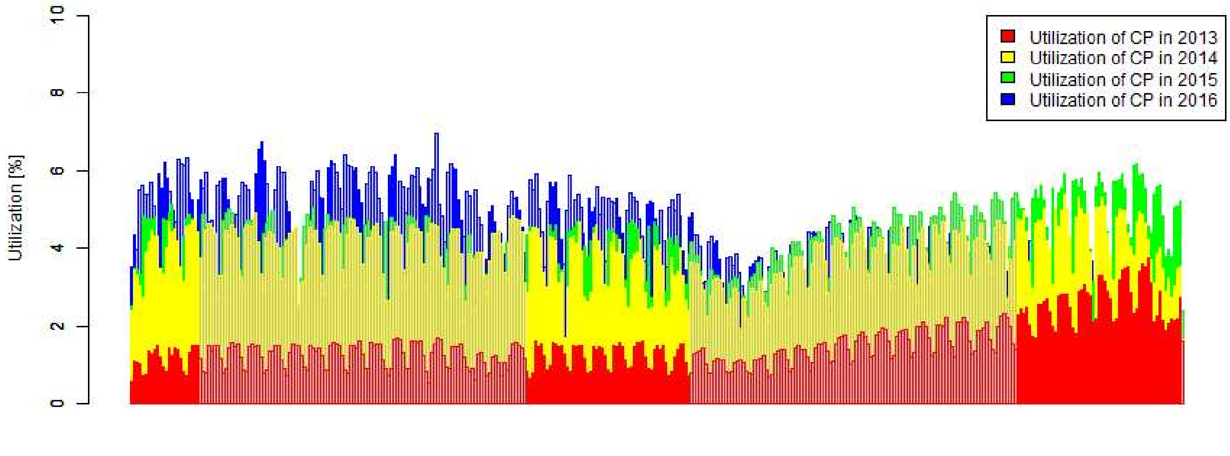
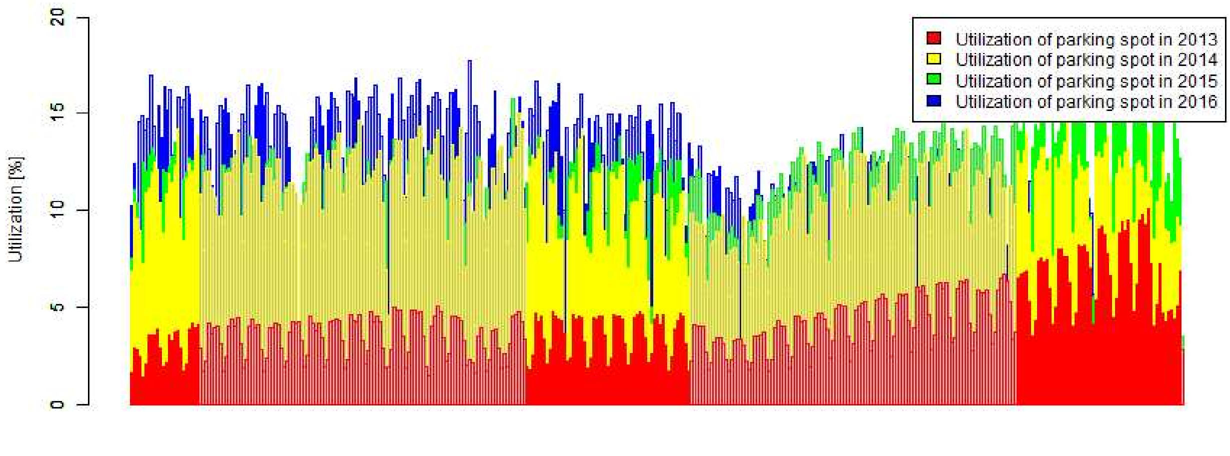



Fig. 3: Comparison of a) Aerial-distance based clustering; and b) Driving-distance based clustering




a)




b)

Fig. 4: Daily utilizations of the a) charging service; and b) parking service for the period 2013-2016

Tab. 3: Average charging and parking utilizations through the period 2013-2016

| Year | Average Utilization of CS [%] | Average utilization of parking spot [%] |
|---|---|---|
| **2013** | 1.50 | 4.12 |
| **2014** | 3.72 | 11.20 |
| **2015** | 4.10 | 11.82 |
| **2016** | 4.62 | 12.71 |

Although the dataset for 2016 consists of 1,765 CSes, some of them have more than one charging plug (CP). Total number of charging plugs in the dataset is 2,922 (for year 2016). The top 500 CPs (~17%) based on utilization are involved with around 65% of all charging sessions through the year and the rest of CPs have negative impact on an overall average utilization of CPs in Netherlands. That can be seen on histogram that represents number of charging plugs per certain utilization level (Figure 5).

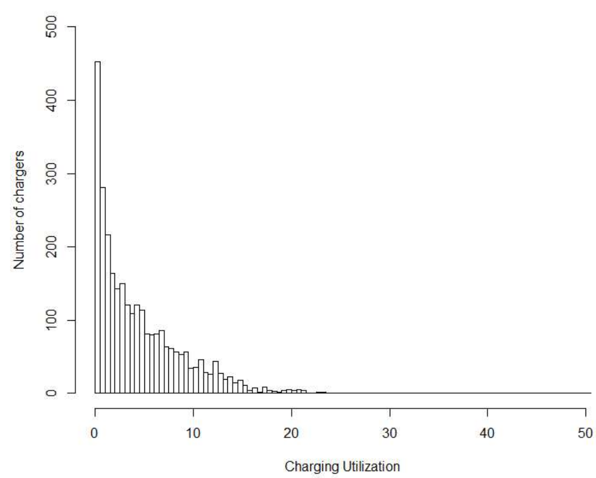



Fig. 5: Number of chargers per utilization level

Figure 6 depicts average charger utilization for 2016. Utilization is calculated per the day of the year (from Monday to Sunday). It can be observed that average utilization for all chargers in 2016 is around 4.50%, while for the top 500 charging plugs utilization is around 13%. Five most utilized chargers have average utilization of 33% while parking spaces associated with them have utilization of over 50%. The most utilized chargers are located near big cities and based on our clustering method, clusters around those cities have greater utilization than others.

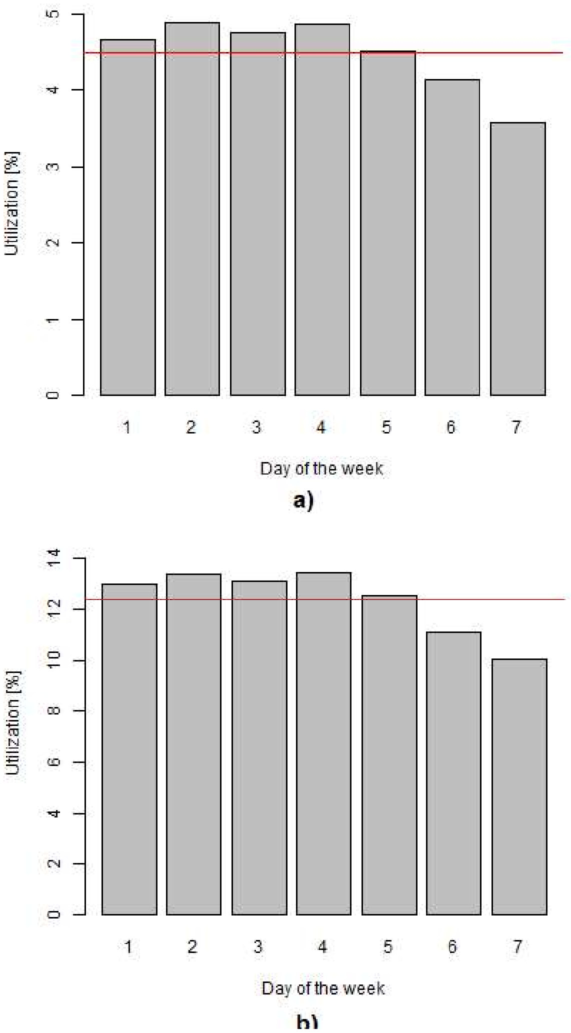



Fig. 6: Comparison of charger utilization for a) all chargers; and b) top 500 chargers

Figure 6 also depicts behaviour that is observable in all years. Utilization is noticeably higher during working days than on weekends. Utilization reaches its peak at the midweek and then starts a near-linear falling. Reasoning behind that is likeliness that people will charge their car during their working day.

Intra-day patterns of usage of chargers and associated parking spots are depicted in Figure 7. Here another pattern can be observed – utilization has two peaks – one around 8 am and the other around 5 pm. These two peaks correspond with (start and end) working hours and match the behaviour of charging a car near working place (8 am) and near home (5 pm). This conclusion is also mentioned in research by Develder et al. (2016), but obtained with different methods. In the same figure utilization of parking spaces can be observed. Utilization of parking spaces does not have peaks, instead it has drops in utilization. Those drops are right before the peaks in utilization of chargers and represent the time of traveling *to* and *from* working place. Previously described pattern can be noticed for all days of the week except weekend (Figure 8). Weekend days have specific distribution of utilization, opposite from working days, with only one peak around 2 pm.

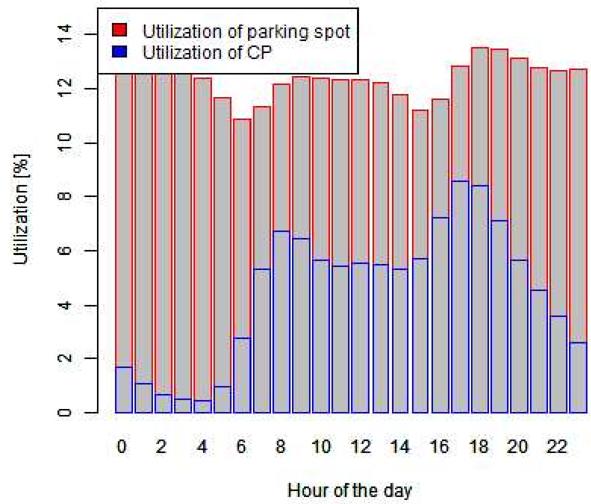



Fig. 7: Comparison of charger utilization and parking spot utilization per hour of the day (weekday)

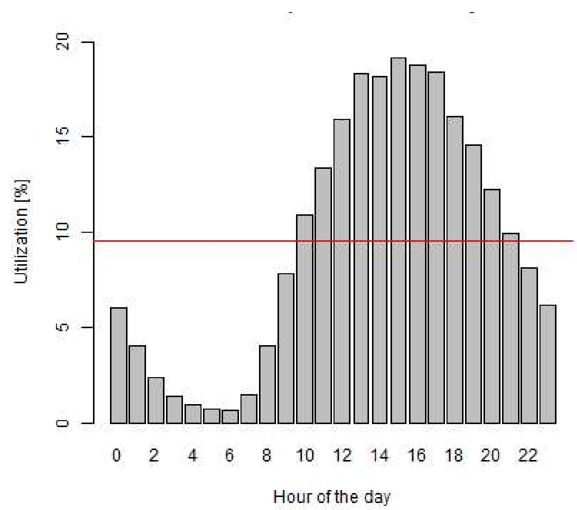



Fig. 8: Utilization (per hour of the day) of chargers on weekends

If utilization of chargers and parking spaces is observed on a level of a day in a year, there are two noticeable drops, the first one and more significant from July to September and the second one in November and December. That pattern corresponds with standard vacation periods and can be observed in Figure 9. The EVCI framework also enables filtering of the dataset and focusing the analysis on a desired timespan only, what enables statistical analysis of different timespans through the year. For example, it is obvious from Figure 9 that not all months have the

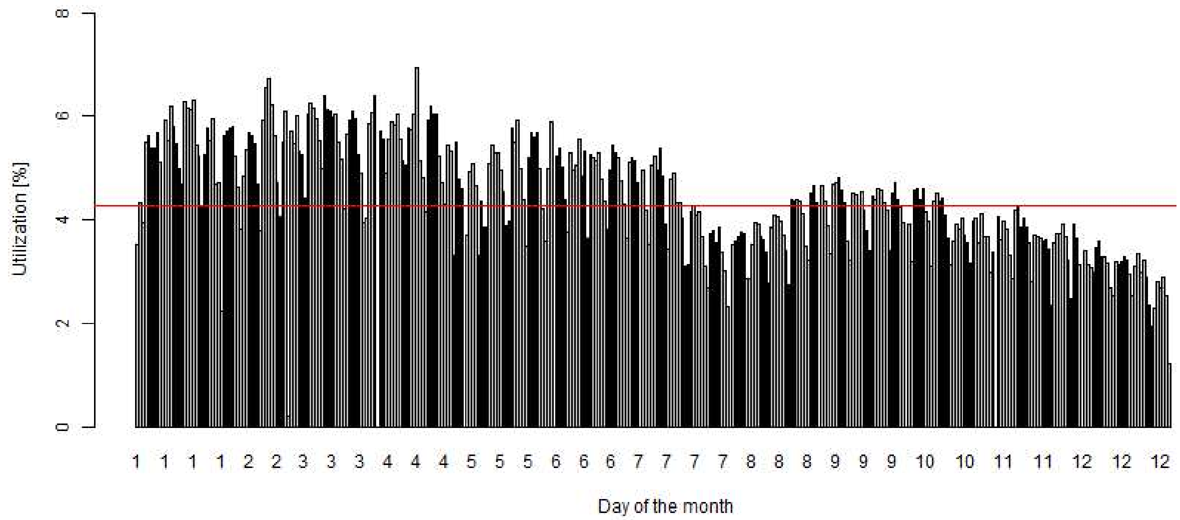



Fig. 9: Yearly charger utilization

same impact on the overall average utilization (*e.g.,* if only first half of the year is observed, average utilization would be significantly higher than utilization observed through a whole year).

### *III.3. Variable representation*

Variable representation is especially interesting, since different variable representations can lead to different conclusions about the dataset and can also be a key factor for better prediction depending on the algorithm selected.

An example of different variable representation is representation of PoIs, which are distributed through 13 categories. If they are represented through their absolute values (*i.e.,* number of PoIs in a certain category) it can be observed how the number of PoIs in a certain category affects utilization (Table 4, *V1*). However, if PoIs are represented as a relative share of PoIs in a certain category taking into account total number of PoIs in all categories, it can be observed how *relation* between PoIs affects utilization (Table 4, *V2*). Finally, PoIs can be represented as a binary value as well (*i.e.,* 1 if a PoI of certain category exist, 0 otherwise; Table 4, *V3*).

Tab. 4: Example of different PoI representation

| **PoI Cat** | **1** | **2** | **3** | **4** | **5** | **6** | **7** |
|---|---|---|---|---|---|---|---|
| **V1** | 0 | 1 | 22 | 12 | 11 | 0 | 3 |
| **V2** | 0.00 | 0.02 | 0.44 | 0.24 | 0.22 | 0.00 | 0.06 |
| **V3** | 0 | 1 | 1 | 1 | 1 | 0 | 1 |

Another example is a reduction of number of variables, what is an important factor which influences computational complexity of certain machine learning algorithms. Namely, our dataset contains information about the start and the end time (*i.e.,* hours and minutes) of a charging session and those four variables were combined to create the fifth variable – *category of the day*. In that way, a day is divided in four categories (*i.e., morning*, *afternoon*, *evening*, and *night*) based on the time span of the session.

### *III.4. Forecasting of utilization*

For accurate prediction of the EVCI framework, selection of appropriate machine learning algorithm is a crucial part of the Model Builder framework module. Machine learning is used to predict utilization of charging stations and parking spots based on historical data (*i.e.,* ground truth) and new input variables (*e.g.,* new available charger for deployment, changed number of PoIs in a cluster, *etc.*). The algorithm used in the EVCI framework implementation version reported in this paper is the *Multiple Linear Regression (MLR).* The MLR algorithm was applied on different time resolutions as follows:

- Day of the year (*e.g.,* every day in a year);
- Day of the week (*e.g.,* every Monday of a year);
- Hour of the day of the week (*e.g.,* 8 am for every Monday of a year);
- Hour of the day (*e.g.,* 8 am for each day through a year); and
- Category of the day (*i.e., morning*, *afternoon*, *evening*, and *night*) of the week in a month of a year.

All time intervals have 17 common variables (*i.e., 13 + 1 PoI categories*, *identificator of a charger cluster*, *number of different EV owners' cards*, and *number of charging plugs in the cluster*) and different variables for specifying time interval. Since all categories of PoIs have high correlation with each other, the new variable was introduced to the dataset – *sumPoI* – which corresponds with the absolute number of all PoIs in the cluster. Although there are 17 variables and time control variables, since time control variables are treated as categorical values together with the identification of cluster, we have many more variables in the model (*e.g.,* if variable is the hour of the day, with the conversion to categorical value this one variable becomes 24 variables: *hour1*, *hour2*, *etc.*).

As it can be seen in Table 5, the MLR algorithm explains around 30% of variance, which is a good result for the *out-of-the-box* algorithm. In the

Table 5 there is also information about significance codes for variables. Based on those values, we can determine how each variable influenced the utilization of chargers in a cluster (*e.g., month4* has less influence on the prediction in comparison to *month7*, if the referent month is *month3*). Based on the significance level, some variables can be excluded from the presented model (but not from alternative models since variables in interaction with each other can carry valuable information). Prediction results presented in Table 5 are based on time interval of *hour of the day of the month on weekend and weekdays*.

Tab. 5 Results of Multiple Linear Regression prediction algorithm

*formula = hourly_actual_charging_time ~ hour_day + isWeekend + month + numCards + numPlug + csid + cluster, data = DataCombined2CS*

| Variable | p-value |
|---|---|
| hour_day_i iϵ[0,23] | *** (hour_day18 **) |
| month_j jϵ[1,12] | *** (month1&2 *) |
| cluster_k kϵ[1, 622] | *** |
| isWeekend | *** |
| numCards | *** |
| numPlug | *** |
| competitorCS | *** |
| Signif. codes | 0 '***', 0.001 '**', 0.01 '*', 0.1 '.', 1 ' ' |
| Residual standard error | 0.06495 |
| Multiple R-squared | 0.30650 |

## IV. Results and discussion

This paper presents the EVCI framework that is able to predict utilization of charging stations and parking spots associated to these charging stations based on historical data of charging sessions. The framework itself is generic and flexible. Those characteristics of the framework are especially valuable as they give users certain dose of freedom to manipulate framework to their needs (*e.g.,* user can define timespan for the machine learning (ML) algorithm and can decide on objective function on which decision about optimal placement of charging stations is made). As the final output of the framework, the user is prompted with descending list of top ten clusters where placement of a new charging station would result in a minimal decrease in overall utilization based on optimization function. Due to its flexibility, the proposed EVCI Framework can also serve as a tool for analysis of the state of the charging infrastructure if some of the chargers are removed.

The Business Analytics and Reporting module of the EVCI Framework is another functionality that is of a great value to the charging infrastructure provider as it grants insights into the current state of its infrastructure and enables analysis from different point of views:

- Utilization of charging stations;
- Utilization of parking spots;
- Overview of utilization of charging plugs (*i.e.,* which ones are more influential in overall utilization);
- Charging habits of users of their infrastructure; and
- Analysis on custom timespans.

Analysis is performed on the data from 2013 to 2016 which enables tracking of evolution of EV-based green transport through years, together with changes of EV owners' behaviour regarding to charging sessions.

Although all modules of the EVCI Framework are implemented, there is still a lot of place for improvement. Since the flexibility of the proposed framework enables quick and easy testing of different algorithms, next step is to test different machine learning algorithms with supporting meta-algorithms (*e.g.,* AdaBoost, Support Vector Machine and Random Forest) and to implement different variable transformations to further improve accuracy. Furthermore, the plan is to acquire new data (*e.g.,* data from other EV charger infrastructure providers, additional contextual data, *etc.*) and to include new variables (*e.g.*, number of EVs that are likely to charge in a certain cluster, information about the speed of the specific charging station, type of the location CS is located in, *etc.*). While the development of the EVCI Framework was based on the data provided by ElaadNL, next step is to test the robustness of the framework on the data provided by other major EV charging infrastructure providers.

Currently, the EVCI Framework is able to answer the proposed research question based on three different objective functions:

(i) <u>minimizing the drop of average charger utilization in a cluster</u>, defined as

$$\min: \frac{\sum_1^{numCharg} Util_{Curr} - \sum_1^{numCharg} Util_{Ext}}{\sum_1^{numCharg} Util_{Current}},$$

where $UtilCurr$ represents current average utilization of all charging stations in a specific cluster before adding another charger, and $UtilExt$ represents the value of the same indicator after EV charging infrastructure is extended (*e.g.,* when more chargers are deployed);

(ii) <u>increasing the number of chargers in a charger unpopulated area</u>; and

(iii) <u>hybrid weighted function taking into account first two objective functions</u>.

Based on the first objective function, to minimize average utilization drop of EV charging infrastructure operated by the ElaadNL on the level of a cluster, the new charger should be added to the *"cluster 525"* (Figure 11: a 3 km radius from the place marked on map) which is located in the fairly populated part of Netherlands (close to 3 large cities: Rotterdam, Hague, and Amsterdam) and currently comprises only four charging stations operated by the ElaadNL. Besides ElaadNL charging stations, in the *"cluster 525"* there are also 9 charging stations of other EV charging infrastructure providers. If another ElaadNL charging station is added, average utilization of the ElaadNL charging infrastructure in the *“cluster 525”* would have a decrease of only 0.3%, which is not an unusual result since that cluster has 97 charging cards (assuming charging cards correspond with the number of EVs, there is around 90 EVs in that cluster).


Fig. 11: Proposed location for the new ElaadNL charging station

## V. Conclusions

The proposed EVCI Framework can be used by EV charging infrastructure providers as the decision support tool which prescribes optimal area to place a new charging station or by governments as a policy development tool which enables them to measure impact of incentives targeting increase of EV owners.

To answer the main research question formed as: “Where should an EV charging infrastructure provider add a new CS?” it is not enough to just properly cluster existing CSes and understand current CSes’ and parking spots’ utilization patterns, but it is important to choose appropriate data as well as the most suitable machine learning algorithm which will result in the highest levels of prediction accuracy for future CSes’ and parking spots’ utilizations.

Therefore, based on real-world data, the proposed framework is able to recommend the optimal location of the new CS with respect to the minimal average charging utilization drop. Ultimately, this information is of a great value for three pillars of sustainability: *People*, *Profit*, and *Planet*.

**Profit** – With the EVCI Framework that can predict utilization for newly placed charging station, EV charging station infrastructure providers could extend their network of chargers while increasing aggregate utilization and, consequently, profit.

**People** – While infrastructure providers would benefit from the proposed framework, consumers would too, since increased number of charging stations would lower the range anxiety and enable more reliable source of energy for EV owners.

**Planet** – With increased reliability and lowered range anxiety, EVs would became more popular and their penetration on the market would increase which is beneficial for the environment, because as stated in the introduction, green transportation would significantly lower the GHG and other pollutants.

## Acknowledgements

The authors acknowledge the support of the research project “Managing Trust and Coordinating Interactions in Smart Networks of People, Machines and Organizations”, funded by the Croatian Science Foundation under project UIP-11-2013-8813. Furthermore, we are thankful to ElaadNL for providing us the data describing their charging infrastructure as well as for their useful comments on the earlier versions of the manuscript and to Oplaadpalen for providing us with API for accessing their data about all chargers in the Netherlands.

## References

Aabrandt, A., Andersen, P. B., Pedersen, A. B., You, S., Poulsen, B., O'Connell, N., & Østergaard, J. (2012, January). Prediction and optimization methods for electric vehicle charging schedules in the EDISON project. In *Innovative Smart Grid Technologies (ISGT), 2012 IEEE PES* (pp. 1-7). IEEE.

Adnan, N., Nordin, S. M., Rahman, I., Vasant, P. M., & Noor, A. (2016). A comprehensive review on theoretical framework-based electric vehicle consumer adoption research. International Journal of Energy Research.

Babic, J., Carvalho, A., Ketter, W., & Podobnik, V. (2017). Electricity Trading Agent for EV-enabled Parking Lots. *Lecture Notes in Business Information Processing*, 271, 35-49.

Bikcora, C., Refa, N., Verheijen, L., & Weiland, S. (2016, October). Prediction of availability and charging rate at charging stations for electric vehicles. In *Probabilistic Methods Applied to Power Systems (PMAPS), 2016 International Conference on* (pp. 1-6). IEEE.

Develder, C., Sadeghianpourhamami, N., Strobbe, M., & Refa, N. (2016, November). Quantifying flexibility in EV charging as DR potential: Analysis of two real-world data sets. In Smart Grid Communications (SmartGridComm), 2016 IEEE International Conference on (pp. 600-605). IEEE.

Kahlen, M., & Ketter, W. (2015, January). Aggregating Electric Cars to Sustainable Virtual Power Plants: The Value of Flexibility in Future Electricity Markets. In *AAAI* (pp. 665-671).

Kayser, M. “A Data-Driven Approach to Modeling Charging Station Utilization Levels by Plug-In Electric Vehicles Type Charging Behavior and Point of Interests”, Master Thesis, Erasmus University, Rotterdam School of Management, Rotterdam, 2016.

Ketter, W., Peters, M., Collins, J., & Gupta, A. (2015). Competitive benchmarking: an IS research approach to address wicked problems with big data and analytics.

Langer, A., & McRae, S. (2013). Fueling Alternatives: Evidence from Naturalistic Driving Data. Working paper.

National Enterprising Netherlands. Figures electric transport. Retrieved February 17, 2017, from http://www.rvo.nl/onderwerpen/duurzaam-ondernemen/energie-en-milieu-innovaties/elektrisch-rijden/stand-van-zaken/cijfers

Neubauer, J., & Wood, E. (2014). The impact of range anxiety and home, workplace, and public charging infrastructure on simulated battery electric vehicle lifetime utility. *Journal of power sources*, 257, 12-20.

Oplaadpalen. Retrieved January 05, 2017, from https://oplaadpalen.nl

Qian, K., Zhou, C., Allan, M., & Yuan, Y. (2011). Modeling of load demand due to EV battery charging in distribution systems. *IEEE Transactions on Power Systems*, 26(2), 802-810.

Saber, A. Y., & Venayagamoorthy, G. K. (2011). Plug-in vehicles and renewable energy sources for cost and emission reductions. *IEEE Transactions on Industrial Electronics*, 58(4), 1229-1238.

Statista. Average speed in Europe's 15 most congested cities Statistic. Retrieved January 05, 2017, from https://www.statista.com/statistics/264703/average-speed-in-europes-15-most-congested-cities

Valogianni, K., Ketter, W., & Collins, J. (2015, May). A multiagent approach to variable-rate electric vehicle charging coordination. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems* (pp. 1131-1139). International Foundation for Autonomous Agents and Multiagent Systems.

Wagner, S., Götzinger, M., & Neumann, D. (2013). Optimal location of charging stations in smart cities: A points of interest based approach.

Yi, Z., & Bauer, P. H. (2016). Optimization models for placement of an energy-aware electric vehicle charging infrastructure. Transportation Research Part E: Logistics and Transportation Review, 91, 227-244.

Xi, X., Sioshansi, R., & Marano, V. (2013). Simulation–optimization model for location of a public electric vehicle charging infrastructure. Transportation Research Part D: Transport and Environment, 22, 60-69.