Data-driven energy demand forecasting for electric vehicle charging infrastructure

Silvia Meddi[4], Sara Cavaglion[1], Tania Cerquitelli[2], Emanuele Manfredi[1], Andrea Regalia[1], Raffaele Menolascino[1] and Guido Zardo[3]

[1]*Accenture S.p.A.*
[2]*Department of Control and Computer Engineering, Politecnico di Torino*
[3]*F2M eSolutions*
[4]*Politecnico di Torino*

Abstract
In the era of Big Data and electric vehicles growth by market, data-driven methodologies assume a crucial role to create valuable information. The focus is on supporting the decision-making process for the development of an accurate charging infrastructure. Forecast analysis allows prediction of energy demand over the network. This supports growing trends with a consequent increase in customer satisfaction. By anticipating potential breakdowns due to infrastructure overloads, maintenance costs are reduced. In this paper, we focus on analyzing charging sessions data together with external data (*weather* and *population information* and *energy/fuel prices*) collected from different sources. The proposed methodology, named GEORGE (enerGy dEmand fOrecasting foR charGing infrastructurE), offers a building-blocks based approach for the monthly energy demand forecasting. The approach is both generalisable and data-specific. We discuss the results of a classification learning approach to predict a belonging range of kwh for a charge point. In particular the most promising model has good performances in predicting high utilization and is more advantageous to support the company's decision-making process. Many possible developments are discussed to improve the prediction.

Keywords
Data Mining, Electric Vehicle, Predictive Model, Applied Data Science, Interpretable Model, Energy Demand

Introduction

In order to ensure transport emissions reduction in Europe and to meet, for example, the *Paris Agreement*, [1], and *Agenda 2030* objectives, [2], many goals need to be achieved over the next twenty to forty years. Two of the 17 goals, set by the Agenda, to be completed by 2030 at European level are: *the fight against climate change* and the creation of *sustainable communities and cities*. These objectives include the following targets. Integration of national policies, strategies and plans against climate change and reduction of negative environmental impact of cities with regard to air quality. Transport sector in general and circulation of cars in particular directly influence greenhouse gas emissions, being responsible of 12% of total emissions at European level, [3]. Thus, the promotion of e-mobility through an interconnected and optimized network of charging stations will help to reduce CO_2. Moreover, the overproduction of energy due to a disproportionate installation of charging stations will decrease, [4]. A relevant player in this scenario is F2M eSolutions that designs innovative technologies to lead the transition to electric mobility. They offer charging solutions and services that will make this change intuitive and seamless. Additionally, with the anticipation of widespread adoption of Electric Vehicles (EVs) in Europe, both public and private developers of charging infrastructure heavily rely on proper data collection and usage to make informed decisions, [5]. Electric mobility is currently a topic of great interest for Europe and for researchers. Many studies focus on how to take advantages from data to support Low-Carbon Road Transport Policies in Europe, [6]. Both the development of charging infrastructures, [7] and the promotion of e-mobility to improve the user experience, [8] depend on an accurate utilization of data. Finding a general method to examine the impact of external factors on the energy demand for charging infrastructure in Italy is currently an unresolved topic in the literature. From a business side, this paper aims to answer to two main necessities. First, support and improve the management of charging stations through a forecast of monthly energy demand. Second, let the client to identify targeted interventions in specific areas. These actions can be supported by interpretable and accessible results that allow instant interaction with data. To reach the goals we used F2M eSolutions' charg-

Published in the Workshop Proceedings of the EDBT/ICDT 2023 Joint Conference (March 28-March 31, 2023, Ioannina, Greece).
s282109@studenti.polito.it (S. Meddi);
sara.cavaglion@accenture.com (S. Cavaglion);
tania.cerquitelli@polito.it (T. Cerquitelli);
emanuele.manfredi@accenture.com (E. Manfredi);
andrea.regalia@accenture.com (A. Regalia);
raffaele.menolascino@accenture.com (R. Menolascino);
guido.zardo@f2m-esolutions.com (G. Zardo)
© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

CEUR-WS.org/Vol-3379/DARLI-AP_2023_7.pdf

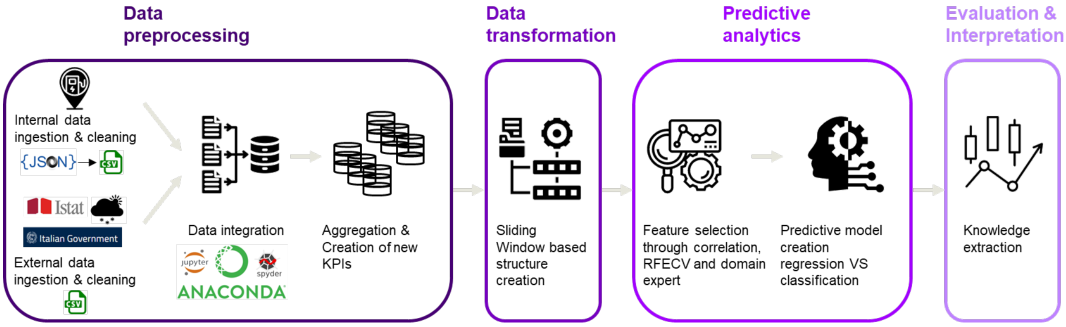

Figure 1: GEORGE Building Blocks.

ing session historical data integrated with external data sources. Data collection and data preprocessing were the most time-consuming activities. The collection goes from the company's internal data (*charge point ID*, *City*, *Country*, *kwh*, ...) to datasets containing *information about the population*, *weather data*, *gas oil/lpg/fuel and energy prices* and *characteristics of the territory*, collected from third parties. Different data science models like *Random Forest*, [9], and *XGBoost*, [10], have been trained to determine the approach that better represents the phenomenon under analysis. In the considered case study, we followed a classification approach as it was the most consistent with business needs: predicting the kwh range of monthly energy demand. In fact, unlike a daily forecast, it provided a general overview of the infrastructure status. This approach showed good results in predicting the observations belonging to the class with the highest demand, which was the major focus from a business side. The final output is thus able to support business decisions in order to achieve a more efficient charging infrastructure usage. The use case on which the project has been developed is the 2022 Italian scenario. The paper is organized as follow. The first section is composed by a brief literature review. It introduces the state-of-art on the promotion and support of e-mobility through data. Section two details the methodology, from the preprocessing step to the evaluation of the predictive results. Section three shows the results obtained by using the proposed methodology on Italian use case. Finally, section four summarizes the content of our work, providing conclusions and suggestions for future improvements.

1. Literature Review

Nowadays, collecting vast amounts of data has become a widespread practice in many scenarios. If done properly, it can be a powerful tool that can offer significant benefits to companies. With the growing market of electric vehicles, interconnection and data transfer technologies, many research orientate the analysis on data-centric approaches.

Studies cited in this section can be grouped in four macro-categories: (i) *application of data science to the world of EVs*, (ii) *support for the transition to electric* , (iii) *monitoring of infrastructure utilisation through Key Performance Indicators (KPIs)* and (iv) *study of the exogenous factors that affect energy demand.*

Applications of data science methods to electric scenario already include a wide spectrum of supervised and unsupervised learning approaches. Several works have already been proposed in this context. [5] analyses the economic benefits of applying data-driven models, i.e. data mining, and machine learning techniques, in a business context. Paper [11] provides a comprehensive overview of use cases that link this scenario to data science, through machine learning algorithms. The authors emphasise the scientific interest of the study in the field of e-mobility. A data-driven approach to extract useful information from electric vehicle charging events is also suggested in paper [12]. In this case, a framework was developed to characterise the demand for electric vehicle charging in a specific geographical area. A key step in this methodology is the one of data cleaning and formatting that is defined specifically for dealing with EVs charging data.

Investing into realistic technological solutions, empowering citizens and aligning action in key areas, i.e. industrial policy, ensure a fair transition to electric. The authors of [13] and [8] focus their studies on predicting charging point occupancy, to support this shift. In [13] the aim is to understand whether the city of San Diego has sufficient charging points to meet the energy demand of EVs through a quantitative and qualitative analysis. In [8], the goal is to support users in planning their charging processes. In particular, they provide a double approach: classification to predict individual charging point occupancy and regression to predict overall charging station occupancy both in public and a workplace site. [14] compares performances of two different approaches. The first one based the installation of new charging points on a request by an electric vehicle driver. The second one chooses for the placement of a CP near strategic locations, considering the decision of a local government. Results show that not one rollout strategy is favorable over the other. Moreover, the best strategy needs to be chosen according to municipal objectives, the maturity of the market and the technologies available.

The large amount of data collected brings with it the need to understand how to extract the most meaningful knowledge. Monitoring properly constructed KPIs usually allows to achieve the goal. [15] presents a study developed using data from the vast public charging infrastructure of the Dutch metropolitan area. The researchers want to identify different charging patterns between five areas and relevant related KPIs. Thanks to forecasting and simulation models, they answer significant questions like where, when and what type of new charge points should be installed. In the development of the study, [7], the authors propose a web-based dashboards to explain particular well- or ill-performing charging stations. The platform aims to support the projected growth of electric mobility through the extraction of relevant knowledge. Performances of the existing charging infrastructure, measured by KPIs developed, drive the know-how.

Many studies confirm that exogenous factors influence charging behaviors. Just to name a few, [6] presents a data processing platform TEMA (Transport tEchnology and Mobility Assessment) designed for supporting EU transport policies through big data. This study shows the implementation of a method capable of managing a significant amount of data from various sources. Thanks to data-driven model, TEMA is able to recognize subtle connections and hidden patterns, performing customized analyses. Many governments started to base their charging network definition on data-driven roll-out strategies. The study [1] identifies and interprets the most impactful characteristics that are correlated with energy consumption. Authors offer useful perspectives on what data need to be utilized to create prediction models and to guide the planning and implementation of charging infrastructure.

Literature shows that several prediction models enable relevant knowledge to be extracted from the data. Most of the studies are very objective-specific and do not allow generalization. Main contributions brought by our study are:

1. bring the analysis to Italian level. The growing circulation of electric vehicles in Italy allows to start analysing the Italian scenario. Until now this panorama is little mentioned in literature;
2. introduce a *Sliding Window based approach* to align the input structure to the final need;
3. propose a general approach that can be expanded, thanks to the integration of external data sources, to investigate the influence of different factors on future behaviors.

2. Methodology

Here we present the GEORGE methodology, whose building-blocks representation is shown in Figure 1. It combines a solid theoretical background with a necessity to solve real business needs. It uses historical data to monitor charging stations and develop forecasting models to predict the monthly energy demand. It consists of a KDD process (Knowledge Discovery process from Data) adapted to business needs: support more efficient development and utilization of the charging infrastructure.
GEORGE consists of four building blocks:

- *data preprocessing* step to clean and merge data from different sources;
- *data transformation* block to adapt the input data structure to the final purpose;
- *predictive analytics* step to derive the most suitable descriptive model to perform accurate predictions;
- *evaluation and interpretation of results* to assess the goodness of the model through specific metrics.

In the following sub-sections, we reported a detailed description of each building block.

2.1. Data Preprocessing

Data preprocessing usually is the first step in the KDD pipeline because prepares data for analysis in the most suitable way. For internal data, the json files recorded for each charging sessions were stored in AWS (Amazon Web Services) s3 bucket. Data was transformed into tables via Athena, an analytics service of AWS, downloaded and saved locally as csv files. The external data was extracted from open source data base in csv format. All input data was processed through python code using Anaconda environment (e.g. Spyder). GEORGE preprocessing was customized in three specific steps: (i) *Data Ingestion and Cleaning*, (ii) *Data Integration from different sources and Alignment to the same granularity*, (iii) *Data Aggregation and Creation of new KPIs.*

The objective of the Data Preprocessing was to have two macro categories of features:

Monthly Based Features (named *Type 1* in the following) that are information available on a monthly basis and include: *charging session information*, *weather data* and *energy, gas oil, fuel and lpg prices.*

Social-Economic and Geographical Features (named *Type 2* in the following) refer to economic and social aspects of the population and to characteristics of the territory. This set includes, for example: *average age of the population*, *employment rate* and *population density.*

Since environmental factors can influence energy demand, we included external data in addition to charging behavior information. The comparison with domain experts led to the following considerations. External temperatures can affect performance of electric vehicles batteries and therefore the number of charging sessions. Moreover, energy prices have direct influence on the demand. On the other hand, since an electric car owner also owns a traditional vehicle, we decide to include gas oil, fuel and lpg prices. Social-economic and geographical factors are able to model the wealth of the population and therefore, indirectly, the ownership of electric vehicles.

Data ingestion for *Type 1* features originated from various sources:

- F2M eSolutions, for charge points and charging sessions data;
- Meteo.it, for weather information;
- Italian Government web site, for prices trends.

and for this reason they had different granularities:

- geographical level
  - state, region, city, longitude and latitude for charge points and charging session;
  - region and province for weather information;
  - state for prices;
- time level
  - year/month/day hour:minutes:seconds for charging sessions;
  - year/month/day for weather information;
  - month for prices.

The step of data cleaning was managed in a specific way for each data source. In presence of internal anomaly, e.g., transmission or estimation error, the corresponding data was removed.
In case of external information not available, different strategies were implemented like replace the missing values with the closest geographical and temporal information.

During data integration, different data sources were aligned to the same granularity using supporting datasets, e.g., for geographical mapping.

Once, the collections had the same granularity we proceeded with the aggregation and definition of new KPIs. For each couple $(i, j) = (chargepoint, month)$, with at least one charging session, we computed the following metrics through a monthly level aggregation: *total number of sessions*, *total session time*, *kwh provided* and *main statistics* for weather values.

From them we computed the following additional indicators, *average session time* and *charge point Occupation Rate*:

$$\textit{ave. sess. time [min]}_{ij} = \frac{\text{tot. session time}_{ij}\ [\text{min}]}{\text{tot. \#of sessions}_{ij}}$$

$$\textit{OR [\%]}_{ij} = \frac{\text{tot. session time}_{ij}[\text{min}]}{\text{tot. \#of days}_j \times 24 \times 60[\text{min}]} \times 100$$

From *Type 2* features prospective, the preprocessing was easier. Data ingestion source included only Italian National Institute of Statistics (ISTAT) and data granularities were at Italian region and annual level. It was not possible to decrease the aggregation data level and therefore this set of information was simply added to the dataset, skipping step (ii) and (iii).

2.2. Data transformation

Data transformation step can be essential for the success of the entire KDD process and it is typically very project-specific.

In order to forecast energy consumption using historical data, we decided for a *Sliding Window based approach.* This method uses a moving window to analyse sequential data and predict future values over a time period. The approach is justified by the capacity of historical trends to affect current needs. We defined two variables, one to identify the *time window* width, *X*, and the other one to define the time horizon for the prediction, or *time forecast window*, *Y*.

Data transformation process consisted of four steps to modify dataset structure, plus a preliminary step for the encoding of all nominal variables.

Step 1 We divided dataset into smaller ones each with information of historical data and prediction window, e.g. $X + Y$ months. X months, then named as $M - X, \ldots,$ $M - 1$, are for the construction of the time window while the last month, $M0$, refers to the time horizon.

Step 2 Not all charge points had data about the entire time window. In absence of useful information to reconstruct the lack of charging session data, we decided to fill missing information with zeros. Thus, we modelled in the same way, a hypothetical failure of charge points and lack of charging sessions due to other reasons.

Step 3 Separately, each subdataset has been transformed to obtain the *Sliding Window based structure.* Figure 2 the configuration with a time window of two months and a time horizon of one month.

Step 4 Finally, all sub-datasets were merged together, resulting in a single dataset where rows contained predictors for $X + Y$ consecutive months.

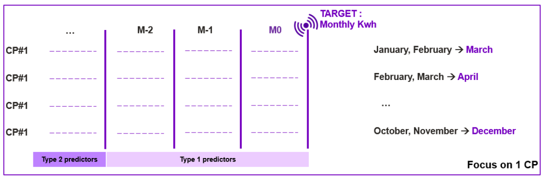

Figure 2: *Sliding Window based approach* for one Charge Point (CP) with $X = 2$ months of time window and $Y = 1$ month of time horizon. *Type 1* predictors represent monthly based features while *Type 2* predictors social-economic and geographical features.

2.3. Predictive analytics

To obtain the data-driven model that best represented the analysed phenomenon, GEORGE required the validation of different algorithms. Therefore, the optimal model was the one with the best combination of performance and ability to provide results.

Monthly energy forecasting could be approached using two different strategies: punctual or categorical value forecasting. While the first one is more specific, it would require a large amount of data to obtain accurate models. Therefore, a multi-class classification task would be of wide interest on several case studies where, especially initially, data might be limited. In fact, from a business side was definitely more functional to have a range of kwh rather than a point value.

We defined classes based on different levels of *percentiles*, a statistical approach that grants balanced classes. This procedure allowed the identification of different amounts of load for the charging infrastructure.

GEORGE integrated various forecasting models using *Sliding Window based structure* described above. Methods chosen are tree-based: *Random Forest* (for details, see [9]) and *XGBoost* (eXtreme Gradient Boosting), see [10]. These are two examples of ensemble learning techniques. In particular, *Random Forest* combines multiple decision tree models to improve general performance. One of the advantages is that this technique has higher accuracy than single decision trees. Additionally it is interpretable, robust to noise and outliers. Then, the statistical framework of *XGBoost* casts boosting as a numerical optimization problem. The objective is to minimize model loss by adding weak learners using a stochastic gradient descent-like procedure.

Models operated on historical data, corresponding to predictors of *Type 1* and *Type 2* for $M - X, \ldots, M - 1$ and $M0$, to forecast the energy demand for $M0$, which changes for each subset of charge points as shown in Figure 2.

Then, through feature selection phase, conducted by *correlation analysis*, *Recursive Feature Elimination with Cross Validation* (*RFECV*) and domain expert support, only the optimal set of predictors was identified to give rise to the best model. Usually, selecting a restricted set of top predictors helps to reduce noise in the model and makes result interpretation more straightforward and effective for business experts.

2.4. Evaluation and Interpretation

Model evaluation is an essential part to understand reliability of prediction against real value.

In this study, identifying charging points that had a growing trend and whose demand might overload the network was the focus of the analysis. Particularly, we were interested on avoiding misclassification of charge points that belonged to high demand class. Otherwise this underestimation of the load might lead to potential risk of breakage in the infrastructure. In this sense, we were primarily interested in reaching high values for metrics of this class, named H in the following.

To extract the best combination of model-purpose GEORGE integrated stratified-cross validation technique and, in case of limited data, the leave-one-out method. The latter technique consists of using a subset of data for training the model and one to validate it, applying then the learning process for each subset, [16]. For *LOOCV* (*Leave One Out Cross Validation*) the test set consists of one observation.

For the classification task, chosen for business need, model performance was estimated based on misclassification. Below the metrics selected to evaluate multi-class classification models on class *H*, [17]:

Precision

$$p_H = \frac{\text{\#of obs. correctly classified in class } H}{\text{total \#of obs. assigned to class } H}.$$

Recall

$$r_H = \frac{\text{\#of obs. correctly classified in class } H}{\text{total \#of observations belonging to class } H}$$

F1 score for class H is the harmonic mean of the precision and recall

$$F1_H = \frac{2 * p_H * r_H}{p_H + r_H}$$

The interpretation of model performances allowed to re-evaluate some previous steps such as feature selection.

The trade off between results and characteristics of models (such as interpretability) determined the choice of the final model to be applied to real time data.

In the following section we present the preliminary results of the optimal model for a use case in Italy.

3. Preliminary results

An accurate manipulation of the huge amount of data that are available today is essential in order to make informed decisions. In this scenario the Energy Demand Forecaster for Italian charging infrastructure can assume a relevant role.

The goal was to create a model, based on temporal aggregation of input data, able to predict the monthly energy demand for single charge point. At the same time we were interested on understanding which factors assumed relevant role in the prediction and their interpretation.

The business value is to support informed decisions to help the network optimization. Moreover, adjusting the load of the charging infrastructure can help in avoiding overloads and breakages.

3.1. Data

In this section we analyse the results of a case study in Italy for which we had data collected from April to October 2022. On the territory the highest percentages were 20% in Piedmont and Lombardy and 10% in Lazio and Emilia-Romagna.

Dataset contained data such as spatial, temporal and energy with different granularities: state, region, city, longitude and latitude for the geographical level and year/-month/day hour:minutes:seconds for charging sessions for the time level.

Principal information recorded in the internal data were: *charging session duration*, *geographical localization*, *kwh provided during each sessions*. Some *charge points characteristics* like if it was subject to any restrictions, i.e. was located in a parking station or it was not open 24/7, were also included in the dataset.

On a monthly basis the mean of kwh provided was about 14.000 kwh for a mean of 500 session per month.

3.2. Data Preprocessing

To obtain the best results from predictive model, we had to understand and structure our data precisely. We started with data manipulations steps.

We exploited cases and limit setting, obtaining more than 3.000 charge points with at least one session and about 10.000 charging sessions. Through the aggregation and transformation steps we structured the dataset to develop the forecasting model. After the evaluation of different widths for the *time window* X and with the support of domain experts we set $X = 2$ while for the *forecast window* $Y = 1$. This confirms that energy demand is linked to historical data by a short-term relationship.

The final dataset was thus composed by predictors for three consecutive months of all the charge points. Two months populated the data history while the *kwh* of the last one was identified as target variable.

3.3. Predictive analytics and evaluation

Since different algorithms were integrated into GEORGE we performed various experiments, leveraging LOOCV, to compare performances.

We recall that, from a business side, it was more functional a classification approach that predicted the range of kwh for each charge point rather than a punctual value.

For each classification algorithm, we evaluated the number of classes, the best subset of features, the size of time window and the impact of hyperparameters. The best model-purpose combination was obtained for the Random Forest with 3 classes, 37 predictors summarized in Figure 3, a 2-month data history and the following hyperparameters:

$$\begin{aligned}
&\text{n_estimators} = 250\\
&\text{max_depth} = 109\\
&\text{max_features} = 18\\
&\text{min_samples_leaf} = 9\\
&\text{min_samples_split} = 7
\end{aligned}$$

Where *n_estimators* is the number of trees in the forest, *max_depth* is the the maximum depth of each tree, *max_features* is the maximum number of features to consider when splitting a node, *min_samples_leaf* is the the minimum number of samples required in a node to be at a leaf and *min_samples_split* is the minimum number of observations in any given node in order to split it.

The model showed high *precision*, *recall* and *F1 score* in the third (or H) class, resulting thus more performing on forecasting high usage:

$$\begin{aligned}
p_3 &= 0.66\\
r_3 &= 0.66\\
F1_3 &= 0.65
\end{aligned}$$

Type 2 predictors, such as *population density*, having annual granularity at Italian region level, were information too general to affect a specific prediction. For this reason, these kind of predictors do not result in the table, Figure 3.

Type 1 predictors for M-2	Type 1 predictors for M-1	Type 1 predictors for M0
price_automotive gas oil_M-2	price_automotive gas oil_M-1	*holydays&workdays*_M0
price_energy(€/kWh)_M-2	price_energy(€/kWh)_M-1	
price_euro-super95_M-2	price_euro-super95_M-1	
price_lpg_M-2	price_lpg_M-1	
*holydays&workdays*_M-2	*holydays&workdays*_M-1	
kwh_M-2	kwh_M-1	
numberofsessions_M-2	numberofsessions_M-1	
occupationrate_cp(%)_M-2	occupationrate_cp(%)_M-1	
workdays_M-2	workdays_M-1	
*weather data*_M-2	*weather data*_M-1	
...	...	

Figure 3: Best subset of predictors for classification task.

Figure 4 shows rank of the most relevant features for the prediction based on decrease in *Gini's impurity index* [1]. It confirms the strong link between the Energy Demand Forcasting and trend of previous months as well the *occupancy rate* and *charging sessions duration*.

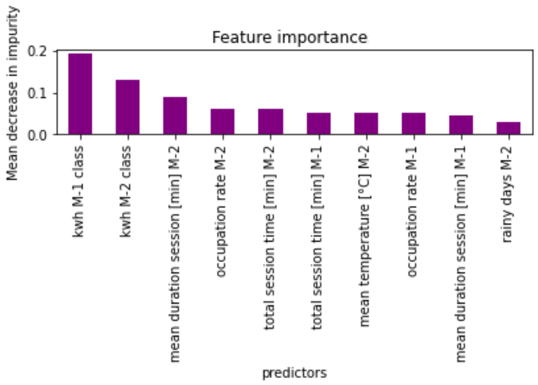

Figure 4: Top 10 important features for Random Forest based on impurity.

4. Discussion

Through GEORGE methodology, based on a solid theoretical background and motivated by concrete business needs, we succeeded in bringing tangible benefits to customer and company. In particular, it made possible the development of an interpretable and generalisable estimator for forecasting the energy demand of the charging infrastructure. The deployment of the model then enables the methodology to be applied to everyday scenario. The real time data stream is processed by GEORGE which extracts relevant patterns. The results are summarised in a customised dashboard that can monitor the state of the network and guide the operators' interventions based on informed knowledge. Figure 5 summarizes key steps of the process on real-time data stream. This can not only improve the management of the charging network, but also the user experience of electric vehicle owners.

[1] Gini's impurity index calculates each feature importance as the sum over the number of splits (accross all tress) that include the feature, proportionately to the number of samples it splits.

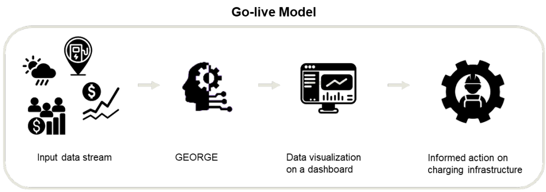

Figure 5: Go-live Model, GEORGE implementation on real-time data stream.

For the specific use case, the biggest effort was on the prediction of class with the highest values of kwh. Correct classifications allow to prevent potential breakages due to an overload of the infrastructure.

Although it is a preliminary work, it is promising because it shows good results for the prediction of the third class. In addition, thanks to the intuitive interpretation, results can be easily shown in an interactive dashboard which can display the history and the forecast for each charge point in navigable map.

Applying GEORGE to Italian's charging infrastructure we can support business decisions with success. Among various benefits, the more relevant are:

- network monitoring based on continuous collection of data;
- promotion of underutilized areas to exploit the infrastructure;
- load monitoring preventing potential breakages and consequent spread of moneys for maintenance;
- implementation of prompt actions guided by predictive analysis;
- support of growing trend in high demand areas, increasing customer satisfaction.

Although GEORGE is a general-purpose methodology, the feature modeling step can not be fully automated. This phase must be modeled according to the final goals, with specific and constant support from domain experts. In addition, understanding which features affect the prediction is an essential aspect to target informed decisions, and it needs conscientious supervision.

GEORGE works with different type of data and this allows future developments of the analysis exploring aspects of the phenomenon as deep and specific as the business needs. It is possible to analyse the influence of population behaviors and characteristics. On the other hand, deeply understanding the impact of weather on battery performance and its consequence on charging demand assume an important role. In this sense, we can leverage GEORGE to address different use cases optimally, adjusting the blocks of the process based on new needs.

Together with the evolution of an electric reality, research progresses and so future directions of this study can be explored.

Possible improvements are:

1. develop an interactive dashboard to monitor the network and support informed decision through the visualization of both historical and forecast trends;
2. increase data history and quality. All the models that we applied are data driven so more quality data will definitely let the model to understand better the relationship between input and output. Decrease granularity, i.e., for the *Type 2* predictors including information at municipality level, will help to explicate the dependence of energy demand on social-economics habits. Considering other type of external data may help to discover more hidden patterns, i.e., traffic information and electric vehicles growth by market for sure impact the forecast of energy demand. The first may support distribution of load through the infrastructure based on pick hours. The latter introduces dependencies on past and, possibly, future sales trend;
3. implement a hierarchical approach that includes a three-class classification model and a regression model for each class to also have a point prediction for each charge point.

In conclusion, improving the dataset is desirable to achieve a prediction based on a larger time forecast window, from a month to 3/6 months, to support the optimization of charging points installation plans.

Acknowledgments

This work was partially supported by

- the European Union under the Italian National Recovery and Resilience Plan (NRRP) of NextGenerationEU, partnership on "Telecommunications of the Future" (PE00000001 - program "RESTART").
- the grant "National Centre for HPC, Big Data and Quantum Computing," CN000013 (approved under the M42C Call for Proposals - Investment 1.4 - Notice"National Centers" - D.D. No. 3138 of 16.12.2021, admitted for funding by MUR Decree No. 1031 of 17.06.2022).

References

[1] M. Straka, R. Carvalho, G. V. D. Poel, L. Buzna, Analysis of energy consumption at slow charging infrastructure for electric vehicles, IEEE Access 9 (2021) 53885–53901. doi:10.1109/ACCESS.2021.3071180.

[2] A. per la coesione territoriale, Agenda 2030 per lo sviluppo sostenibile, 2015. URL: https://www.agenziacoesione.gov.it/comunicazione/agenda-2030-per-lo-sviluppo-sostenibile/.

[3] C. A. European Union, Co2 emission performance standards for cars and vans, 2020. URL: https://climate.ec.europa.eu/eu-action/transport-emissions/road-transport-reducing-co2-emissions-vehicles/co2-emission-performance-standards-cars-and-vans_en.

[4] U. N. E. Programme, Supporting the global shift to electric mobility, 2021. URL: https://www.unep.org/explore-topics/transport/what-we-do/electric-mobility/supporting-global-shift-electric-mobility.

[5] J. Sung, M. Troilo, N. Howarth, Better energy efficiency policy with digital tools, 2021. URL: https://www.iea.org/articles/better-energy-efficiency-policy-with-digital-tools.

[6] M. De Gennaro, E. Paffumi, G. Martini, Big data for supporting low-carbon road transport policies in europe: Applications, challenges and opportunities, Big Data Research 6 (2016) 11–25. URL: https://doi.org/10.1016/j.bdr.2016.04.003. doi:10.1016/j.bdr.2016.04.003.

[7] S. Maase, X. Dilrosun, M. Kooi, R. Van den Hoed, Performance of electric vehicle charging infrastructure: Development of an assessment platform based on charging data, World Electric Vehicle Journal 9 (2018). URL: https://www.mdpi.com/2032-6653/9/2/25. doi:10.3390/wevj9020025.

[8] A. Ostermann, Y. Fabel, K. Ouan, H. Koo, Forecasting charging point occupancy using supervised learning algorithms, Energies (2022). URL: https://doi.org/10.3390/en15093409. doi:10.3390/en15093409.

[9] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning. Data Mining, Inference, and Prediction, Springer, New York, NY, 2009.

[10] T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, Association for Computing Machinery, New York, NY, USA, 2016, p. 785–794. URL: https://doi.org/10.1145/2939672.2939785. doi:10.1145/2939672.2939785.

[11] D. Pevec, J. Babic, V. Podobnik, Electric vehicles: A data science perspective review, Electronics 8 (2019). URL: https://www.mdpi.com/2079-9292/8/10/1190. doi:10.3390/electronics8101190.

[12] E. Xydas, C. Marmaras, L. M. Cipcigan, N. Jenkins, S. Carroll, M. Barker, A data-driven approach for characterising the charging demand of electric vehicles: A uk case study, Applied Energy 162 (2016) 763–771. URL: https://www.sciencedirect.com/science/article/pii/S0306261915013938. doi:https://doi.org/10.1016/j.apenergy.2015.10.151.

[13] S. Viswanathan, J. Appel, L. Chang, I. V. Man, R. Saba, A. Gamel, Development of an assessment model for predicting public electric vehicle charging stations, European Transport Research Review 10 (2018) 1–15.

[14] J. Helmus, J. Spoelstra, N. Refa, M. Lees, R. van den Hoed, Assessment of public charging infrastructure push and pull rollout strategies: The case of the netherlands, 2018. doi:10.1016/j.enpol.2018.06.011.

[15] R. Wolbertus, R. van den Hoed, S. Maase, Benchmarking charging infrastructure utilization, World Electric Vehicle Journal 8 (2016) 754–771. URL: https://doi.org/10.3390/wevj8040754. doi:10.3390/wevj8040754.

[16] C. Sammut, G. I. Webb (Eds.), Leave-One-Out Cross-Validation, Springer US, Boston, MA, 2010, pp. 600–601. URL: https://doi.org/10.1007/978-0-387-30164-8_469. doi:10.1007/978-0-387-30164-8_469.

[17] M. Grandini, E. Bagli, G. Visani, Metrics for multi-class classification: an overview, 2020. URL: https://arxiv.org/abs/2008.05756. doi:10.48550/ARXIV.2008.05756.