Predicting EV parking behaviour in shared premises

Vinicius Monteiro de Lira[1], Fabiano Pallonetto[2], Lorenzo Gabrielli[1] and Chiara Renso[1]

[1]Institute of Information Science and Technologies, CNR, Pisa, Italy
[2]School of Business, Maynooth University, Kildare, Ireland

Abstract
The global electric car sales continued to exceed the expectations climbing to over 3 millions and reaching a market share of over 4%. However, uncertainty of generation caused by higher penetration of renewable energies and the advent of Electrical Vehicles (EV) with their additional electricity demand could cause strains to the power system, both at distribution and transmission levels. The present work fits this context in supporting charging optimization for EV in parking premises assuming a incumbent high penetration of EVs in the system. We propose a methodology to predict an estimation of the parking duration in shared parking premises. The final objective is estimating the energy requirement of a specific parking lot, evaluate optimal EVs charging schedule and integrate the scheduling into a smart controller. We formalize the prediction problem as a supervised machine learning task to predict the duration of the parking event before the car leaves the slot. We test the proposed approach in a combination of datasets from 2 different campus facilities in Italy and Brazil. The overall results of the models shows an higher accuracy compared to a statistical analysis based on frequency, indicating a viable route for the development of accurate predictors for sharing parking premises energy management systems.

Keywords
Electrical Vehicles, Parking Prediction, Machine Learning

1. Introduction

The advent of Electrical Vehicles (EV) are in increasing spreading in our society. According to MCkinsley report[1] in our society EV sales rose 65 percent from 2017 to 2018 and Europe has seen the strongest growth in EVs. According to the report "The European Union's new emissions standard—95 grams of carbon dioxide per kilometer for passenger cars—could also boost EV sales because it stipulates that 95 percent of the fleet must meet this standard in 2020 and 100 percent in 2021". A race for larger batteries among manufacturer is leading the current EV technology, and going forward it appears that as batteries technology improve, they are going to replace motor fuel vehicles. The concerns as we move to EVs is that, firstly, there will not be enough charge points to meet consumer demand and, secondly, this additional load on the electricity grid will cause partial and total failure of specific electrical plant due to overloading. The present work fits this context supporting optimization for EV charging and assuming a incumbent high penetration of EVs in the system. We propose an approach to predict an estimation of the parking duration in shared parking premises. This is essential for estimating the energy requirement of a specific parking lot, evaluate optimal EVs charging schedule and integrate the scheduling into a smart controller.

The specific behaviour of parking lots of campuses refereed to EV charge is peculiar since it substantially differs from the general parking lots available in the streets. In campus-like facilities (Universities, large industries, etc) we can observe regular patterns of parking behaviour that mainly include staff working hours besides a part of other visitor [1]. This can be an advantage when trying to predict general behavioral patterns of parking habits and thus reach an optimal recharge plan for EVs.

The current work can become part of an overall design of a smart charging energy management system to optimally integrate the distributed energy systems and EVs into the power grid by developing a parking prediction module to estimate the vehicles' parking time using machine learning algorithms.

Given this context, the specific objective of this work is to predict the duration of each *parking event* in a campus-like parking lot, where the *parking event* is the moment of the actual parking action of a car in a slot. In other words, at the moment the car is parked in a slot, we predict how long this car will stay parked so that the change can be optimized. This will allow the energy management system to decide when to start the actual charge based on the prediction. For example, assuming several cars arrive at 9 AM, if we don't know how much the car stay parked and therefore we start charging immediately, this will cause a peak of electricity demand. On the contrary, if we predict that a given car stay parked e.g. 8 hours, then we

Proceedings of the Workshop on Big Mobility Data Analytics (BMDA) co-located with EDBT/ICDT 2023 Joint Conference (March 28-31, 2023), Ioannina, Greece
viniciusmonteirodelira@isti.cnr.it (V. M. d. Lira); fabiano.pallonetto@mu.ie (F. Pallonetto); lorenzo.gabrielli@isti.cnr.it (L. Gabrielli); chiara.renso@isti.cnr.it (C. Renso)

[1]https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/mckinsey-electric-vehicle-index-europe-cushions-a-global-plunge-in-ev-sales#

can delay the charge of this car to a later moment trying to flatten the peaks. We formalize the prediction problem as a supervised machine learning task that, given a parking event at a given time, tries to predict the duration of the parking event. We structure our experiments inspired by the research questions: **RQ**: *How accurately a supervised machine learning approach can predict the duration of a parking event in a campus-like parking lot?* We experiment different algorithms and features combination into 5 datasets from 2 different campus facilities in Italy and Brazil. We show that using both contextual and time of the day features, the overall results of the models shows an higher accuracy compared to a statistical analysis based on frequency, indicating a viable route for the development of accurate predictors for sharing parking premises energy management systems.

Structure of the paper follows. Section 2 discusses some related works an how this approach differentiates from them. Section 3 introduces the problem definition and clarifies the prediction problem formulation. Section 4 reports the details of the experimental evaluation. Finally, Section 5 draws the conclusions and envisage some future works.

2. Related Works

Most of the recent works in the literature focuses on predicting which parking lot will be free at the arrival of the car. This is motivated by the challenge of finding a parking space in urban areas. It has been reported[2] that 30% of traffic congestion is caused by travelling for finding parking spaces, bringing unnecessary energy consumption and environmental pollution. Works in this area include the off-street (parking slots in private areas) and in-street variants (slots in the streets). A pioneering paper by [2] proposes the real-time availability forecast algorithm to predict parking facility availability in real time using combined current (on-line) and historical information. This work uses an algorithm operating with mixed real and simulated information.

A recent paper by Zeng et al ([3]) proposes a hybrid model that stacks gated recurrent unit (GRU) and long-short term memory (LSTM). The GRU-LSTM model combines LSTM's advantage in prediction accuracy and GRU's advantage in prediction efficiency. Furthermore, similar to us, it uses multi factors, including occupancy, weather conditions and holiday, as input to predict parking availability.

In paper [4] authors develop a prediction model based on Naive Bayes and machine learning methods like decision tree, random forest, and regression analysis for building the prediction model of parking occupancy and therefore predict the subsequent parking availability combining a matching-based allocation strategy to assign users to selected parking spaces.

Deep learning to predict parking occupancy is proposed by many papers in the literature. Paper [5] adopts a deep learning model for predicting block-level parking occupancy 30 min in advance in paper. The model takes multi-source data as input, e.g., parking, traffic and weather.

Paper [6] proposes a Convolutional Neural Network (CNN) model for block-level parking occupancy prediction extracting spatial relations of traffic flow combined with a LSTM autoencoder to capture temporal correlations. Clustering is also considered in the Clustering Augmented Learning Method (CALM) to learn deep feature representations of spatio-temporal data obtained using the proposed embedding.

A deep learning approach is also proposed by [7], called Du-Parking. This approach models temporal closeness, period and current general influence employing long short-term memory (LSTM) to model the temporal closeness and period. This approach learns to dynamically aggregate the output to estimate the final parking availability of given parking lot.

Compared to the above approaches the novelty of our method lies in the fact that we do not aim at predicting the next free slot neither to suggest the driver where to park. On the contrary, we predict how long the car will stay parked in a given slot and this prediction is computed at the specific time the car start the parking, which is a different problem. While the first problem requires to train a learning system to predict which slots - in a given area at a given time - will be free, this problem, given a specific slot and a specific starting time of the parking of a vehicle, predicts how long the car will stay parked. The energy management system can therefore allocate the energy to the parking slot charging station based on this prediction: charging later in time vehicles which are predicted to stay longer and accelerate the charging for cars which are predicted to stay shorter.

3. The parking duration prediction problem

The objective of our approach is to exploit historical data on parking usage and additional contextual data like weather conditions and parking lot occupancy levels, to predict the duration of a parking slot occupancy. Differently from many state of the art approaches that want to predict if a giving parking lot will be free in a next period of time [2, 4], here we focus on the prediction of the temporal duration of the occupancy of a car in a slot. We recall that our approach, to be suitably integrated with an Energy Management System, focuses on specific

[2] https://www.accessmagazine.org/spring-2011/free-parking-free-markets/

parking context that we call of shared premises (e.g. parking lots of universities, workplaces, supermarkets, etc), not focusing on fee-based street parking. Overall, our approach can be applied to any parking environment where there is a tendency for the car to stay parked a minimum amount of time and where the electrical charge system of the parking lots can be integrated into the controller of the Energy Management System. We also recall that our approach is driver-profile agnostic due to privacy reasons.

It is worth noticing that the parking behaviour in a campus-like facility reflects a different parking behaviour compared to fare-based streets parking lots since here the duration is expected to be longer than on street parking.Furthermore, the premises have usually a controlled access and the energy management system can therefore optimise the electricity supply based on the parking occupancy, while is not always true in fare-based street parking.

Given a parking area, a *car parking event* represents an event where a driver parks at a given timestamp in one of the available slots. The vehicle stays parked for a certain temporal duration until it leaves the slot. It is assumed that the vehicle can be charged while parked. The charging time can start as soon as the car arrives, or can start later on, or again, can start, interrupt and start again.

Definition 1 (Parking Slot). *Given a parking area in shared premises, we define a parking slot s as a tuple $s =< id, x, y >$, where id represents the parking slot identifier, and x and y represent its spatial coordinates in the parking area. A parking slot is the actual place where drivers park their cars. In our application scenario, each slot can be equipped with a charge station where the car can be recharged. The set of all the parking slots form the parking area.*

Definition 2 (Car Parking Event). *We define a car parking event e as a tuple $e =< s_{id}, t_{start}, d_e >$, where s_{id} represents the parking slot identifier where the car is parked, t_{start} represents the timestamp indicating when a car has started the parking and d_e is the temporal duration of the car park until it leaves the slot.*

We want to predict the parking duration d_e of a car parking event e, given a $slot_{id}$ and the parking event starting time t_{start}. This prediction is modelled as a classification problem where the objective is to assign, for each car parking event e, a class representing the predicted duration interval. More formally, we have the following definition of the problem.

Definition 3 (Parking Duration Prediction Problem). *Given a parking event e where it is known the slot identifier s_id and the start time t_{start} but not duration d_e, we want to define a function $f(s\_id, t_{start}) = c$ where the class c represents a temporal interval such that $d_e \subseteq c$.*

We can observe that our target variable c represents ordinal categories. An ordinal variable is a categorical variable, where there is a clear ordering of the categories. For example, our variable could assume ordinal categories like: short, medium or long duration. In the next section, we introduce the details of the Machine Learning (ML) approach to solve the Car Parking Duration Prediction Problem.

We propose to use supervised machine learning approaches to predict the parking duration based on an historical dataset of car parking events and contextual features.

The learning task is based on a three types of features: *single event-related*, *spatial* and *contextual* features. The event-related features represent the features that we can extract directly from the sets of parking events like the time of the parking event or the weather conditions. The spatial features are based on the location of the parking slots inside the car parking area, while the contextual features representing the occupancy of the different zones of the parking area. We explain how we have extracted some spatial and contextual features that are used in our predictive models. All these features are combined to feed the proposed supervised machine learning algorithms (Section 4).

Recalling Definition 2, a car parking event is defined by the tuple $e =< s_{id}, t_{start}, d_e >$ and given s_{id} and t_{start} we want to predict the temporal duration d_e. From the timestamp t_{start}, we derive three features: the day of week dw, hour of the day h, and the minutes m rounded to 5 minutes. The motivation of these temporal features is to enable the predictive model to learn the correlation between the time when the car parks and the relative parking temporal duration. We also include in this category of features the weather condition wr at the moment of the car parking event starts, t_{start}, using this as extra information to feed the predictive models.

Many studies have been made toward the understanding of parking behavior and the mechanism of people's parking decisions [8, 9, 10, 11]. It has been observed that some spatial aspects can bias the occupancy of a parking lot and choose the parking areas close to the destination but also investigated the impact of the trees in parking lots.

Motivated by these aspects, our approach focus on the spatial distribution of the parking slots. For this reason, we split the whole parking lot into smaller areas using different clustering approaches. Then, we include these spatial features in our predictive models to learn if a parking area can correlate with the slot occupancy duration.

More formally, given a set of parking slots $(s_0, ..., s_n) \in$

S, we use the spatial coordinates of each s_i to create the spatial clusters $(a_1, a_2, .., a_k)$ of parking slots, where $k \leq n$. We have used two clustering algorithms for this task: DBScan ([12]) and K-Means ([13]). Thus, when training our predictive models over the the dataset of historical car parking events, we add as input feature a representation of the cluster where the parking slot s_i belongs to.

Another aspect that we investigate for the parking duration prediction is the context. In our case the context is represented by the status of occupancy of the slots in the spatial clusters and the relationship of this occupancy with the duration of a given parking event.

Specifically, we want to discover if the occupancy status of an area (e.g 100%, means totally full, while 0% totally empty) where a driver parks, has a correlation with the parking duration.

The contextual features therefore represent the status of occupancy of the different areas (i.e. the spatial clusters) of the parking. In other words, to predict the duration of a given car parking event d_e, we also consider as input feature the level of occupancy of the spatial clusters $(a_1, a_2, .., a_k)$ at the time of the event.

More formally, we define a function called *occupancy* that given a $slot_{id}$ and the timestamp t_{start} of a parking event e, creates a vector $(o_{a_1}, o_{a_2}, .., o_{a_k})$ by computing the occupancy level of all parking spatial clusters $(a_1, a_2, .., a_k)$ at time t_{start}. The occupancy status o_{a_j} of each spatial cluster a_j is basically the ratio between the number of occupied slots at the timestamp t_{start} and the number of slots in that cluster.

In Figure 1, we illustrate the spatial and contextual features. At the top, we can see images from three different car parking lots. In the middle, the dots represent the pixel coordinate of the parking slots. The colors of each dots is a representation of the spatial features which indicates the cluster (sub-area) of each parking slot.

In the bottom we illustrate the contextual occupancy features. The color represents the spatial cluster of each slot. At the bottom, we have the occupancy status in percent of the spatial clusters a_1, a_2, a_3, and a_4 at a given time t.

To summarize, in this section we introduced three new categories of features: event-related, spatial and contextual. The event-related features are extracted from the parking event, the spatial features are computed by using cluster techniques over the spatial distribution of the parking slots while the contextual features are obtained by computing the occupancy status of the spatial clusters.

The overall idea is to investigate how to train the predictive models using different information that might have a predictive power on the parking duration. In the next section we detail the experimental setting and results on exploiting these features in a machine learning task for predicting the parking duration of a given event.

4. Experimental evaluation

In this section we evaluated the proposed approach for predicting the parking duration by exploiting historical parking data.

The research question driving our experiments is the following:

RQ: *How accurately a supervised machine learning approach can predict the duration of a parking event in a campus-like parking lot?*

This research question guides our first experiments. Here, we compare the performance results of our machine learning based approach against several baselines. We also investigate different machine learning approaches to tackle this problem as a supervised task: Classification, Ordinal Regression, and Regression. We use different features including the description of the occupancy of the parking lot at the time the parking event starts and the spatial distribution of the parking areas.

4.1. Experimental Setup

Datasets. We selected two public datasets of parking occupancy in campus-like parking lots: PKlot [14] and CNRPark [15]. Both datasets contain the occupancy information detected by video cameras for each slots of parking areas of two academic institutions: the research area of the National Research Council of Pisa[3], in Italy and the parking area of the two Brazilian universities. In both cases the whole parking lot is split in different parking areas with a variable number of parking slots. In both datasets, a car parking event occurs when a car parks in a parking slot of the area. In this case, the event starts at the timestamp of the frame that detects a car in the slot. The car parking event ends at the timestamp of the frame showing: (1) an empty parking slot, or (2) a different car parked in the same slot. The duration of the parking event is then computed as the difference of the timestamps of the two image frames, the start and the end. The CNRPark dataset contains images collected from November 2015 to February 2016 for a total of 23 monitored days. The parking lot has been monitored by 9 cameras covering the parking area of the Pisa National Research Council south parking area. The meteorological conditions at the moment of the frame capture has also been collected. This dataset contains a total of 4081 frames and 144,965 photos. In Table 1 we depict an excerpt of the CNRPark raw data.

The change of cars in the monitored slots is also detected and these statistics are reported in Table 2 for the CNR dataset. We see the day of the week, the number of changed cars (the end of a parking event) and the total number of days for which we have actual images. We

[3]http://www.area.pi.cnr.it

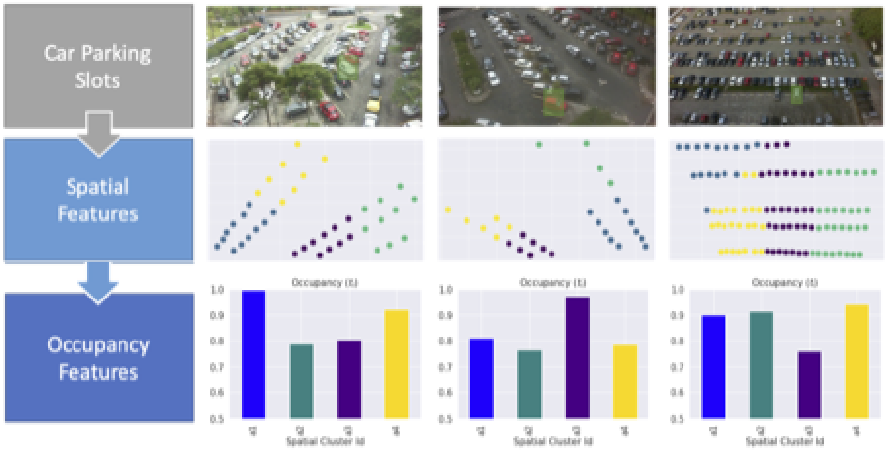

Figure 1: Illustration of the spatial features using the k-means algorithm to split the parking lot into four spatial clusters and the occupancy clusters.

Date	Time	Slot	Occupancy	Weather
12/11/2015	08:15	275	free	SUNNY
12/11/2015	08:45	275	free	SUNNY
12/11/2015	09:15	275	busy	SUNNY
12/11/2015	09:45	275	busy	SUNNY

Table 1
An extract of the CNR parking dataset

Weekday	Change of cars	Monitored Days
Monday	371	2
Tuesday	419	2
Wednesday	392	2
Thursday	1005	5
Friday	1084	6
Saturday	36	4
Sunday	26	2

Table 2
Number of change car events in the CNR dataset

also notice, as expected, how the number of car changes during the week end is very low.

The PKlot dataset contains the occupancy information for each slot of the parking areas of two academic institutions: (1) the Federal University of Parana (UFPR) and (2) the Pontifical Catholic University of Parana (PUCPR), both located in Curitiba, Brazil. The dataset includes a total of three different parking lots represented by PUCPR, UFPR04, and UFPR05. The occupancy information is detected by a number of cameras taking images of the parking slots and detecting the change of the car or the slot becoming empty. This dataset contains 12.417 images captured in three different parking areas with different weather conditions for a total of 168 slots in the period between 11 September 2012 and 16 April 2013. Specifically, dataset PUCPR has 100 parking slots, UFPR04 has 28 and UFPR05 has 45 slots. PKLot is larger than CNRPark and contains images spanning across months.

Data Cleaning. A detailed analysis on the image frames reveals the presence of missing data for some hour of the day (e.g. due to a broken device or during night hours due to the lack of infrared vision). When a parking event is starting or ending during the missing temporal interval, it has been flagged as partial and filtered out to avoid training inconsistencies. Additionally, we filter out the days when the number of detected slots is lower than 50% of the total slots. In total, after the data cleaning CNRPark counted 3552 parking events, PUCPR 4291 parking events, UFPR04 1204 parking events and UFPR04 2148 parking events.

Weekday	CNRPark	PUCPR	UFPR05	UFPR04
Monday	2	3	4	2
Tuesday	2	8	6	5
Wednesday	2	5	4	4
Thursday	5	5	4	4
Friday	6	7	5	5
Saturday	4	5	5	4
Sunday	2	4	2	1

Table 3
Number of monitored days in the parking lots datasets

Target Classes. We have considered the following 3 classes for the predictive variable (i.e. the car parking event duration) with discrete values in minutes: $Short \leq 60$, $60 < Mid \leq 240$, $Long > 240$; Table 4 show the normalized distribution of the car parking events.

	Train				Test			
	Short	Mid	Long	Entropy	Short	Mid	Long	Entropy
CNRPark	0.09	0.31	0.60	0.81	0.13	0.36	0.51	0.89
PUCPR	0.41	0.28	0.31	0.99	0.35	0.32	0.33	1.00
UFPR04	0.52	0.28	0.20	0.92	0.44	0.28	0.28	0.98
UFPR05	0.44	0.26	0.30	0.98	0.54	0.32	0.14	0.89

Table 4
Normalized distribution of the car parking events in the train and the test datasets. For each dataset we also report the entropy calculated according to the frequency values of the classes

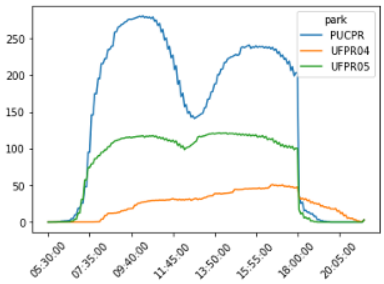

Figure 2: Distribution of occupancy of the parking lots of PKLot (CNR, omittted due to lack of space, has similar figures)

Training Approaches. Given the ordinal characteristic of our target variable, we have explored three supervised approaches to train the predictive models over the training set. For each approach, the best model is selected taking into account the average results over the 5 folds of validation. These approaches are: (a) Classification: the training is performed without taking into account the order of the classes and the selected model is the one with highest *micro-fscore; (b) Regression: the training is executed to reduce the* mean square error (MAE) of the predicted values, therefore the model with lowest error is selected.

Algorithms. For the Classification and Regression tasks we used the following algorithms: Random Forest (RF), XGBoosting (XGB), AdaBoosting (AB), Logistic Regression (LR) and Support Vector Machine (SVM). To compute the spatial features, we have used the K-means and the DBScan clustering algorithms. For all algorithms, we used the implementation available in the scikit-learn library[4].

Features. The following features are extracted and used to feed the ML algorithms. The event-related features include hour of the day h, time stamp minutes m, day of week dw, slot id s, and weather condition wr; the spatial features include the spatial cluster id spt; the occupancy features include the spatial cluster occupancy ocy. We use different feature combinations to train the models: (1) Single event-related feature where we train the model using only one event-related feature; (2) All event-related features together where we train the model using all single event-related features at once. We refer to all when we use all the event-related features to train the ML model. For both cases, we perform two further combinations: using and not using the spatial and occupancy features to feed the models.

Hyper parameters. We use a grid search to tune the hyperparameters of the algorithms [16]. Specifically: For XGB, AB and RF, we vary the number of trees in the range of {50, 100, 150}, while maximum tree depth vary in the ranges of {2,3, or until all leaves are pure}, respectively; For SVM, we use the RBF kernel with γ varying in {0.0001, 0.001, 0.01}. For the LR, we have used two different class weight parameter {balanced and uniform}, while the multi class parameter changing between {auto, ovr (for binary classification)}. For the K-means, the k varies in the ranges of {2, 3,4,5,6}. While for the DBScan ranges are {50, 75, 100, 125, 150} and {2,3,4} for the ϵ and minimum sample, respectively.

Baselines. To be able to evaluate the performance of our approach we have used the following baselines: (a) Random: randomly choose a class; (b) Longest Class: always select the longest interval; (c) Shortest Interval: always choose the shortest interval; (d) Majority Class: always choose the class with highest frequency in the training data. For regression, we compare with the (e) Linear Regression (LN). Naive Bayes and Linear Regression are both simple ML models with high bias. They are used here as baselines given their easy interpretation.

ML model training process. For each dataset, we split the car parking events into train and test with 0.8 and 0.2 ratio respectively without shuffle the data. To avoid data leakage, we ordered the car parking events using their timestamps before split. When training the models on the training data, we use a stratified cross-validation with 5 folds. After the training, for each algorithm, the best configuration of hyper-parameters is used to retrain the model using the whole training data and then assess its performance now using the test set.

[4]https://scikit-learn.org/

Evaluation metrics. To evaluate the experiment results we have used the following measures: micro f1-score ($F1_{micro}$), macro f1-score ($F1_{macro}$) and mean absolute error (MAE). We recall that the F1 score is a weighted average of the precision and recall where best value is 1 and worst is 0. The micro f1-score is a metric where we compute an F1 score counting the total true positives, false negatives and false positives. The macro f1-score is a metric that treats all classes equally, then it does not take label imbalance into account. Indeed, the macro-average computes the metric independently for each class and then take the average, hence treating all classes equally, whereas the micro-average will aggregate the contributions of all classes to compute the average metric. These measures give some clues about the precision and recall of the models on predicting the true positives. By using the MAE we want to have a more interpretative measure of our regression models since it computes the average error of the predictions values ($\hat{y}_i$) compared to the real values (y_i). For all experiments, we consider the MAE obtained over the test set as comparison criteria between the models.

4.2. RQ: Accuracy of ML in predicting parking duration

In this section we address our research question - studying the accuracy of our car parking event duration prediction models. At this first study, we analyse the performance of each ML approach ($Classification$ and $Regression$) when predicting parking events duration. Table 5 report the MAE, micro f1-score and macro f1-score of the models. The MAE was used as comparison criteria to select the best models. For each dataset and ML approach pairs, the table indicate the strongest baseline and report the results of the best ML models for two set of features: (a) using only event-related features, represented as $Alg_{\{h,m,dw,s,wr,all\}}$, having no spatial and occupancy features; and (b) using event-related features with spatial and occupancy features, represented as $Alg^{\{spt,ocy\}}$. The table also report the improvement in percentage achieved by the ML models over the baselines. We highlight in bold the best MAE result per dataset.

From the results, we can observe that the ML models overcome the baselines in all the datasets, for all the training approaches. Specially the ensemble trees models (RF and XGB) show the best results in most of the training approaches with XGB showing the best performance. For all datasets, when using a $Regression$ approach, we observe that the most robust baseline is the Linear Regression (LN), whereas for the approaches $Classification$ the strongest baseline is the Gaussian Naive Bayes (GNB). In Table 5 the best MAE performance (0.316) is reached by classification with XGB when predicting over the PUCPR dataset. The best f1_micro and f1_macro are also recorded in the PUCPR dataset using the classification task. Moreover, we observe that the use of the spatial and occupancy features in most of the cases (16 out of 24) has improved the performance of the ML models.

Altogether, these analysis show an consistent advantage in the use of low bias ML models such as the XGB to predict parking event duration due to the implicit randomness and non-linearity of such events.

Both classification and regression algorithms produce a similar performance for long term parking; however, classification is more accurate on the short term forecasting while regression has an overall lower mean average error in the medium range. The parking prediction module based on classification could provide a better user experience to drivers because accurate identification of short-term parking will force the controller to guarantee a higher energy share to short-term park events. However, it will reduce the peak shaving capabilities of the parking area. On the other hand, a regression model could facilitate demand response measures because of forecasted parking events shifted towards long parking time.

5. Conclusions and Future Works

The growing penetration of EVs with larger battery size challenges the distribution network's capacity, and it is becoming a threat to the grid's reliability. The use of EV batteries as flexible energy storage opens up several research questions on integrating charging vehicles with an Energy Management System. Such a system requires implementing a parking occupancy prediction module, proposed in this paper, based on historical data to forecasting the parking duration for each parking slot. We evaluated various machine learning algorithms across four different parking datasets to predict parking behaviour in this context. Future works include the improvement of the current performance results with finer prediction intervals and more classification options.

Acknowledgements

The work is supported by the ERA-NET Smart Energy System, Sustainable Energy Authority Ireland and Italian Ministry of Research with project EVCHIP N. ENSGPLUS-REGSYS18_00013. and GA 19/RDD/579-EVCHIP.

References

[1] U. Duda-Wiertel, A. Szarata, The analysis of transport-related behaviours of drivers in highly

Dataset	Approach	Baseline	Algorithm	MAE	F1$_{micro}$	F1$_{macro}$
CNR	Classification	GNB$_h$	RF$_h$	0.345 (+5.8%)	**0.714** (+7.6%)	0.597 (+32.2%)
			XGB$_{all}^{ocy,spt}$	**0.342** (+6.5%)	**0.714** (+7.6%)	**0.606** (+34.2%)
	Regression	LN$_{all}$	SVM$_{all}$	0.417 (+11.6%)	0.696 (+3.3%)	0.586 (+13.8%)
			SVM$_{all}^{ocy}$	0.397 (+13.2%)	0.689 (+2.3%)	0.581 (+13.0%)
PUCPR	Classification	GNB$_h$	XGB$_h$	0.341 (+42.0%)	0.702 (+21.8%)	0.701 (+49.8%)
			XGB$_{all}^{ocy,spt}$	**0.316** (+46.3%)	**0.735** (+27.5%)	**0.73** (+55.9%)
	Regression	LN$_{all}$	RF$_{all}$	0.39 (+43.2%)	0.693 (+74.0%)	0.692 (+132.1%)
			RF$_{all}^{ocy,spt}$	0.4 (+41.8%)	0.692 (+73.7%)	0.691 (+131.6%)
UFPR04	Classification	GNB$_h$	RF$_h$	0.556 (+3.6%)*	0.568 (+3.0%)	0.504 (+9.8%)
			XGB$_{all}^{ocy,spt}$	**0.51** (+11.5%)*	**0.593** (+7.5%)	**0.547** (+19.2%)
	Regression	LN$_{all}$	XGB$_{all}$	0.574 (+9.7%)	0.494 (+9.2%)	0.502 (+14.9%)
			XGB$_{all}^{ocy,spt}$	0.571 (+10.1%)	0.481 (+6.4%)	0.478 (+9.2%)
UFPR05	Classification	GNB$_{all}$	RF$_h$	**0.340** (+16.6%)	**0.702** (+8.2%)	**0.645** (+13.4%)
			XGB$_h^{ocy,spt}$	0.379 (+6.9%)	0.672 (+3.6%)	0.6 (+5.6%)
	Regression	LN$_{all}$	RF$_h$	0.463 (+11.0%)	0.6 (+17.8%)	0.592 (+61.0%)
			XGB$_{all}^{ocy,spt}$	0.456 (+12.4%)	0.593 (+16.4%)	0.571 (+55.4%)

Table 5

Car parking event duration prediction results comparing the top 1 results of each approach. All the improvement against the respective baseline present statistical significant differences on the residuals compared to the best baseline (Dependent t-test for paired samples' test with 95% confidence interval), except those indicated with *.

occupied paid parking zones., Advances in Transportation Studies 47 (2019) 35 – 48.

[2] F. Caicedo, C. Blazquez, P. Miranda, Prediction of parking space availability in real time, Expert Systems with Applications 39 (2012) 7281 – 7290.

[3] C. Zeng, C. Ma, K. Wang, Z. Cui, Parking occupancy prediction method based on multi factors and stacked gru-lstm, IEEE Access 10 (2022).

[4] E. H.-C. Lu, C.-H. Liao, Prediction-based parking allocation framework in urban environments, International Journal of Geographical Information Science 34 (2020) 1873–1901.

[5] Z. Mei, W. Zhang, L. Zhang, D. Wang, Real-time multistep prediction of public parking spaces based on fourier transform–least squares support vector regression, Journal of Intelligent Transportation Systems 24 (2020) 68–80.

[6] S. S. Ghosal, A. Bani, A. Amrouss, I. El Hallaoui, A deep learning approach to predict parking occupancy using cluster augmented learning method, in: 2019 International Conference on Data Mining Workshops (ICDMW), 2019, pp. 581–586.

[7] Y. Rong, Z. Xu, R. Yan, X. Ma, Du-parking: Spatio-temporal big data tells you realtime parking availability, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2018, p. 646–654.

[8] F. Zong, P. Yu, J. Tang, X. Sun, Understanding parking decisions with structural equation modeling, Physica A: Statistical Mechanics and its Applications 523 (2019) 408 – 417.

[9] G. Antolín, A. Ibeas, B. Alonso, L. dell'Olio, Modelling parking behaviour considering users heterogeneities, Transport Policy 67 (2018) 23–30.

[10] X. Ma, X. Sun, Y. He, Y. Chen, Parking choice behavior investigation: A case study at beijing lama temple, Procedia-Social and Behavioral Sciences 96 (2013) 2635–2642.

[11] D. D. Milošević, I. V. Bajšanski, S. M. Savić, Influence of changing trees locations on thermal comfort on street parking lot and footways, Urban forestry & urban greening 23 (2017) 113–124.

[12] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise, in: KDD, AAAI, 1996.

[13] J. B. MacQueen, Some methods for classification and analysis of multivariate observations, in: Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, 1967.

[14] P. R. De Almeida, L. S. Oliveira, A. S. Britto Jr, E. J. Silva Jr, A. L. Koerich, Pklot–a robust dataset for parking lot classification, Expert Systems with Applications 42 (2015) 4937–4949.

[15] G. Amato, F. Carrara, F. Falchi, C. Gennaro, C. Vairo, Car parking occupancy detection using smart camera networks and deep learning, in: IEEE Symposium on Computers and Communication, 2016.

[16] J. S. Bergstra, R. Bardenet, Y. Bengio, B. Kégl, Algorithms for hyper-parameter optimization, in: Advances in neural information processing systems, 2011, pp. 2546–2554.