Expert Systems With Applications 238 (2024) 122264

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Bike sharing and cable car demand forecasting using machine learning and deep learning multivariate time series approaches

César Peláez-Rodríguez *, Jorge Pérez-Aracil, Dušan Fister, Ricardo Torres-López, Sancho Salcedo-Sanz

Department of Signal Processing and Communications, Universidad de Alcalá, Campus Universitario, Ctra. Madrid-Barcelona km 33, 28805 Alcalá de Henares, Spain

ARTICLE INFO

Keywords:
Cities green mobility
Bike sharing demand prediction
Cable car demand prediction
Machine learning
Deep learning

ABSTRACT

In this paper the performance of different Machine Learning and Deep Learning approaches is evaluated in problems related to green mobility in big cities. Specifically, the forecasting of bike sharing demand in Madrid and Barcelona (Spain) is approached, for different prediction time-horizons, and also a problem of cable car demand forecasting in Madrid city. An important number of predictive variables are considered, which are grouped into four different sets (categorical/calendrical, persistence-based, meteorological and, as a novelty of the paper, information about analogue past instances), whose relevance is studied for all cases. A feature selection mechanism is also incorporated in order to improve the prediction accuracy of the proposed algorithms. A total of 12 different multivariate regression techniques are implemented, covering from Machine Learning methods to time-series Deep Learning approaches. Excellent results in all the prediction problems approached are reported. Finally, the consequences of obtaining accurate prediction in these three problem of green mobility in big cities are discussed. In addition, it is studied how the results could be exported to other similar cases in more general urban mobility studies. Novelties of the work include: (1) Addressing the forecast problem of passenger flow on a cable car using ML and DL multivariate techniques; (2) using the demand of analogous past instances as an additional feature to solve the demand prediction problems; and (3) the extraction of global conclusions about feature relevance when addressing a demand forecasting problem in green mobility.

1. Introduction

1.1. Motivation and incitement

Anticipating and characterizing human mobility is fundamental for an efficient organization of cities (Miskolczi, Földes, Munkácsy, & Jászberényi, 2021), with special significance in urban planning (Bassolas et al., 2019; Dokuz, 2021), traffic forecasting (Asencio-Cortés, Florido, Troncoso, & Martínez-Álvarez, 2016; Çolak, Lima, & González, 2016; Kitamura, Chen, Pendyala, & Narayanan, 2000; Van Arem, Kirby, Van Der Vlist, & Whittaker, 1997), accidents prediction and prevention (Charandabi, Gholami, & Bina, 2022; Wang, Pei, Li, & Yao, 2018) or the spread of biological (Eubank et al., 2004; Kraemer et al., 2020) and even electronic viruses (Kleinberg, 2007). The increasing number of public databases where certain human-related services or activities are monitored on a daily or hourly frequency, together with the massive amount of data generated by the use of smartphones (Rawassizadeh, Momeni, Dobbins, Gharibshah, & Pazzani, 2016), have led to a significant proliferation in the study and modeling of human mobility patterns (Guizzardi, Pons, Angelini, & Ranieri, 2021). These studies evidence that human-related activities often exhibit behaviors far from random (Tan et al., 2021), showing patterns with a high degree of regularity (Pappalardo et al., 2015) and predictability (Song, Qu, Blumm, & Barabási, 2010).

Specifically, travel demand prediction has become a rapidly evolving area in recent years, with numerous studies emerging every year to address this issue. The most common approaches include the prediction of shared-transportation-means related metrics such as the number of check-in and check-out in stations (Li, Zheng, Zhang, & Chen, 2015),

The code (and data) in this article has been certified as Reproducible by Code Ocean: (https://codeocean.com/). More information on the Reproducibility Badge Initiative is available at https://www.elsevier.com/physical-sciences-and-engineering/computer-science/journals.

* Corresponding author.
E-mail addresses: cesar.pelaez@uah.es (C. Peláez-Rodríguez), jorge.perezaracil@uah.es (J. Pérez-Aracil), dusan.fister@uah.es (D. Fister), ricardo.torres@edu.uah.es (R. Torres-López), sancho.salcedo@uah.es (S. Salcedo-Sanz).

https://doi.org/10.1016/j.eswa.2023.122264
Received 18 January 2023; Received in revised form 19 October 2023; Accepted 19 October 2023
Available online 21 October 2023

the demand forecasting within a station (Lin, He, & Peeta, 2018), the usage forecasting of dockless bike-sharing (Hua et al., 2020), or the use of optimization methods to find the best inventory (Márquez, Bracho, & Ramirez-Nafarrate, 2021), resource allocation (Xie, Li, Liu, & Tan, 2023) or rebalancing procedure (Benchimol et al., 2011).

1.2. Related work

One of the key common features of problems related to green mobility in cities is that these problems, as all human-related activities, exhibit regular patterns and strong repeating rhythms that are characteristic of hourly, daily, weekly or monthly time labels. Therefore, these problems are usually represented as multivariate timed data, where the patterns exhibited by the target variables are strongly influenced by exogenous variables. The exogenous predictors choice usually constitutes a critical step in the development of a predictive demand forecasting application, being the human mobility specially dependent on: (1) situational and calendrical context: occurrence of major or public holidays, type of day, weekends, holidays, or any hourly, weekly, monthly and yearly seasonal factors (Deville et al., 2014; Sardinha, Finamore, & Henriques, 2021; Sathishkumar, Park, & Cho, 2020; Torres-López et al., 2022); and (2) meteorological context: Weather conditions, including precipitation, humidity temperature or any other environmental event of relevance (Kim, 2018).

In addition, although it is possible to find in the literature studies focusing on spatio-temporal information (Li et al., 2021; Li, Zhu, Kong, Xu and Zhao, 2019; Yang, Heppenstall, Turner, & Comber, 2020) the majority of the approaches, including the one presented in this paper, are focused on time-series based prediction (Collini, Nesi, & Pantaleo, 2021; Lv, Zhi, Sun, & Qi, 2021; Peng et al., 2021; Zhang, Zhuge, Jia, Shi, & Wang, 2021). In this context, in order to exploit the temporal relations among the different variables contained in complex datasets, Machine Learning (ML) and Deep Learning (DL) methods have been widely used to tackle this problem. In Toch, Lerner, Ben-Zion, and Ben-Gal (2019) the analysis of large-scale human mobility by the use of ML techniques is carried out. In Charandabi et al. (2022) a generalized regression neural network is used to predict road accident risk when inputting 22 predictor variables. In Souto and Liebig (2016) both statistical methods, unsupervised learning techniques and feature analysis are applied for anomaly detection on spatial time series for urban traffic applications. In Wang et al. (2018) an algorithm for risky driver recognition based on vehicle speed time series is proposed. Regarding green human mobility, the majority of ML techniques has been applied to problems of bike sharing demand estimation in cities. For example, in Tekouabou et al. (2021) a combination of Internet of Things and ML techniques are proposed to optimize the management of self-service shared bike systems in smarts cities. Specifically, a problem of bike sharing demand estimation in London is tackled in that work. In Sathishkumar et al. (2020) different ML approaches such as Gradient Boosting Machines, Support Vector Regression algorithms (SVR), Boosted Trees and Extreme Gradient Boosting Trees (XGBT) are tested in a problem of bike sharing demand prediction. Predictive variables include data from weather, day of the week and type of the day (holidays, etc.). Excellent prediction results were reported in a problem of bike demand estimation in Seoul (South Korea). In Ve and Cho (2020) RF has been used for predicting the hourly rental bike demand in Seoul (South Korea). The dataset included bike-sharing metrics, meteorological features and date information collected along 12 months with hourly time granularity. In Li, Zhao and Li (2019), Discrete Wavelet Transform (DWT) was used to reduce dimensionality and filter out random errors of the raw time series, then, time series were clustered using k-means based on similarities measured by Dynamic Time Warping (DTW) and prototypes computed using DTW barycenter averaging (DBA). The proposed approaches were applied on a 3-month bike usage dataset acquired on the bike-sharing system of Chicago. In Ashqar et al. (2017) RF and Least-Squares Boosting (LSBoost) were used as univariate regression algorithms, and Partial Least-Squares Regression (PLSR) was applied as a multivariate regression algorithm for modeling the availability of bikes at San Francisco Bay Area Bike Share stations. Finally, in Sathishkumar et al. (2020) different ML methods were evaluated for the hourly bike sharing demand prediction, including both meteorological features, number of bikes rented per hour and date information as predictor variables. In Harikrishnakumar and Nannapaneni (2023) Quantum Bayesian Networks (QBN) are proposed to provide computational speedup in comparison with classical algorithms in a problem of bike sharing demand prediction. Results in data from CitiBike dataset in New York City are reported. In Sun and Lu (2023) the authors study land use, as a crucial role in promoting the bike-sharing demand. The paper analyzes different urban regions and different data sizes, with six machine learning fusion methods, including Inverse Distance Weight, Spline, Kriging, Natural Neighborhood and Trend. Experiments in Beijing city shows that the ML fusion models improve the estimation performance compared with individual interpolation algorithms. In Alzaman, Aljuneidi, and Li (2023) an approach which combines machine learning with supply chain network design is proposed for a joint problem of bike usage and optimizing operations at repair shops. Specifically, the paper analyzes the use of ML algorithms, such as neural networks, decision-tree-based regression, K-nearest neighbor, support vectors, and ensemble random forest, to predict bike usage and repair.

Additionally, DL approaches have been increasingly employed due to their capability of extracting deeper relations among temporal variables, as not only the information at any precise instant is taken into account, but also pieces of information from previous steps (Torres, Hadjout, Sebaa, Martínez-Álvarez, & Troncoso, 2021). In Wang and Kim (2018), RF, Long Short-Term Memory networks (LSTM) and Gated Recurrent Units networks (GRU) were compared for the short-term prediction of the number of available bikes in different time ranges of 1, 5 and 10 min, in Suzhou, China. In Collini et al. (2021) Bidirectional Long Short-Term Memory networks (Bi-LSTM) were employed to predict the short-term available bikes in the cities of Siena and Pisa (Italy). Recurrent Neural Networks are also employed in Chen et al. (2020) and Lu and Lin (2020) to predict both rental and return demand in real-time. In Ai et al. (2019), convolutional long short-term memory networks (conv-LSTM) were applied to address the spatial and temporal dependences of the problem. The spatio-temporal variables, including number of bicycles in area, distribution uniformity, usage distribution, and time of day as a sequence are considered, in such a way that both the input and the prediction target are spatio-temporal 3D tensors within one end-to-end learning architecture. In Yang, Xie, Ozbay, Ma, and Wang (2018) a DL approach using the convolutional neural networks (CNNs) was proposed to predict the daily bicycle pickups at both city and station levels. Other than the historical records, relevant information like weather was also incorporated in the modeling process. In Li et al. (2023) a recent try to capture spatial–temporal dependency in a problem of bike sharing demand prediction is carried out by using a modified CNN and LSTM algorithm. The idea is to exploit the fact that bicycle usage in neighboring areas might not always be similar, but there are spatial variations that affect cycling activities. On the other hand, it is known that areas that are far apart can be relatively more similar in temporal usage patterns. Thus, the proposal is to exploit these spatio-temporal patterns by using an irregular convolutional Long-Short Term Memory model (IrConv+LSTM) to improve short-term bike sharing demand forecast. Results in dockless bike sharing systems in Singapore, Chicago, Washington, D.C., New York, and London are used to show the good performance of the proposed approach. In Xu, Di, Yang, Chen, and Zhu (2023) a transformer-encoder-based neural process model is proposed to fit the distribution of bike usage in bike sharing systems. This is one of the first works which incorporate transformer models to improve the capability of extracting relevant information and improve the prediction of the number of bike pickups and returns. The approach was tested in CitiBike datasets in New

York City, with good results. In Yang et al. (2023) a hybrid approach composed by a multigraph convolution network (GCN) to model the built environment, a long short-term memory (LSTM) network to extract temporal features, and a fully connected network (FCN) to model weather influence, is proposed for a problem of bike sharing demand prediction. Experiments in Tianjin, China, show the performance of this hybrid DL proposal in the problem of bike sharing demand prediction. In Jia et al. (2023) a graph-based neural network model is proposed, to learn the representation of bike-sharing demand spatial–temporal graph. The model has the ability to use graph-structured data to represent both spatial- and temporal aspects of the problem into consideration. A case study of bike-sharing demand prediction in Nanjing, China, is presented to evaluate the proposed DL algorithm.

This paper is focused on the prediction of green mobility associated with bike-sharing demand and car cable demand problems in two big Spanish cities, Madrid and Barcelona. While there is a growing volume of literature regarding the demand forecasting of bike sharing services, as previously mentioned, very few works have been reported focusing on the prediction of the passenger demand for such a singular transportation mean as the cable car. Although some studies have focused on the transport systems, construction of cable cars (Hoffmann & Liehl, 2005) or on the assessment of time savings and effects of cable car systems on urban mobility (Garsous, Suárez-Alemán, & Serebrisky, 2019), the study of passenger flow in cable car is often been ignored because of the fixed location and limited capacity. However, in big and hyper-connected cities, the number of cable car passengers can reflect the tourist flow of nearby attractions, and can be considered as an indicator to estimate the tourism density and distribution in the city. Few studies are reported concerning the demand forecasting of cable cars: In Lu and Wei (2017) the concept of pan-holidays is introduced, which includes free promotions as a new holiday. Peak index and k-means clustering are employed to estimate the passenger flow of eight cable car routes in Huangshan Scenic Area (China). In Hofer, Haberl, and Fellendorf (2017) and Hofer, Haberl, Fellendorf, Huber, and Fallast (2018) a multimodal transport model was enhanced with the results of coordinated mobility surveys, aiming at estimating the demand of an additional hypothetical transport system, respectively a cable car, in the city of Graz (Austria). In Guzman, Cantillo-Garcia, Arellana, and Sarmiento (2022) users expectations and perceptions of a new cable car in the southern periphery of Bogotá (Colombia) is evaluated by conducting a panel survey before and after the cable car started operations to evaluate the ranking of preferences toward a set of possible benefits of the project. To the authors' knowledge, no contributions have been found in which ML/DL techniques are used to estimate the daily demand of a cable car system.

1.3. Contribution and paper organization

The approach adopted in this paper has involved the demand forecasting of three problems related to green mobility in cities. Specifically, the bike sharing demand in the city of Madrid, Spain (BiciMad public system) and Barcelona, Spain (Bicing-Barcelona system) has been tackled, along with the passenger flow of Madrid cable car (Teleférico). The three problems have been formulated as multivariate time-series-based regression, where different time resolution (daily and hourly) has been employed according to the specific problem. The datasets were expanded with time-series exogenous variables, which can be grouped into 4 different modules, differing from other similar approaches in the literature:

- Meteorological variables: information regarding temperature and rainfall for the same and previous days
- Categorical/Calendrical context variables: variables related to the type of day, month, time, festivities or any other label in this regard.
- Persistence related variables: information related to the autoregressive pattern, and weekly and monthly averages of user demand.
- Analogous instances variables: as a novelty of this work, variables with information about analogous instances to the current prediction were included as predictor variables, these variables take into account similarities between the current prediction and past events.

In addition, an intrinsic problem in this type of prediction work is dealing with incomplete databases or missing information (Bi, He, Xie, & Luo, 2023; Luo, Wu, Wang, Wang, & Meng, 2021). This limits the quality and quantity of exogenous variables that can be added to each problem. It underscores the importance of customizing and rigorously developing the database during the construction and expansion processes, in order to obtain accurate predictive results.

Different ML and DL methods have been applied to the three problems considered, with a total of 12 different regression methods, including shallow techniques such as linear, polynomial, Lasso regression, algorithms based on regression trees (Regression Trees and Random Forests), Support Vector Regression and different architectures of neural networks: Multilayer Perceptron (MLP) and Extreme Learning Machine (ELM), and deep learning models specific to time series treatment such as Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU) and Convolutional Neural Networks (CNN). Regression problems with different prediction time-horizons have been tackled, covering from hourly to daily prediction. A feature selection algorithm based on exhaustive search have been designed and implemented to select the optimum set of variables for each regression model. In addition, issues related to a high percentage of missing instances in some of the time series databases have been addressed. In spite of this, accurate results have been obtained in the three green mobility databases considered. These results show that ML and DL regression techniques are excellent techniques to tackle these green mobility problems. Finally, and constituting an important contribution of this work with respect to other similar works, an exhaustive feature relevance analysis has been performed to assess the importance of each group of variables in each case, allowing us to explain the critical parameters that contribute the most to a successful demand forecasting.

The main contributions of the article are listed as follows: (1) An exhaustive analysis of which sets of features are most relevant for each prediction method has been carried out. As the same methods and the same groups of variables have been used for the three problems considered, it is possible to extract global conclusions about which variables are more important when addressing a demand forecasting problem in green mobility. (2) To the authors' knowledge, this is the first study in which passenger flow on a cable car is predicted using ML and DL multivariate techniques, achieving remarkable results. (3) Although Analogous-based estimating techniques have been used profusely in many regression and prediction problems (mainly in climatology and meteorology Alessandrini, Delle Monache, Sperati and Cervone, 2015; Alessandrini, Delle Monache, Sperati and Nissen, 2015; Salcedo-Sanz et al., 2019; Vanvyve, Delle Monache, Monaghan, & Pinto, 2015), in this paper the demand of analogous past instances has been considered as an additional feature, that is a novelty in the field of demand forecasting in green mobility problems. In particular, two instances are considered to be "analogous" when they have the same day-of-week label, the same holiday label, temperatures within a range of ± 5 Celsius and rainfall values within a range of ± 1 mm/24 h. Therefore, for each sample we consider as predictors both the demand value of the last analogous instance and the average demand of all analogous instances, taking into account only the events occurring before the one being predicted.

The remainder of the paper has been organized as follows: in Section 2 the timed databases used in this paper are described. Section 3 provides a brief overview of the ML and DL regression techniques considered in this work. Section 4 presents the results obtained for each prediction problem, including the analysis on the performance of each method according to each problem. Section 5 discusses about the future lines of work and the possible ways in which the developed system may help to urban mobility problems and closes the paper with some final remarks and conclusions.

2. Datasets and descriptive statistics

Three different databases related to green mobility in big Spanish cities are used in this paper. All of them consist of public access databases that display the number of users of the service provided by the specific owning company, with a daily or hourly resolution. In particular, data corresponding to the number of bicycles rented per day in the city of Madrid (from the Public Madrid Company *BiciMad*), to the number of bicycles rented per hour in the city of Barcelona (offered by the public service *BicingBarcelona*) and to the number of daily users of Madrid's cable car (managed by the Empresa Municipal de Transportes de Madrid, *EMT*) were used. These three databases are described in more detail in Sections 2.1–2.3.

In all cases, these datasets lead to regression problems, where the target variable to predict corresponds to the bike demand (number of bicycles rented per day (BiciMad), or per hour (Bicing Barcelona)) and the cable car demand (number of daily users (Madrid's cable car)). The predictions are carried out one hour and one day ahead, respectively. Fig. 1 shows the time series of each considered dataset. It may be observed that the Madrid's cable car dataset presents discontinuous data, which complicates the prediction problem. In addition, as previously mentioned, it is important to complement datasets with exogenous information in order to improve the prediction efficiency. Input features can be grouped into four different sets according to its concerning information: (1) Meteorological data (relative to temperature and precipitation) for the location of each dataset have been used. They were extracted from ERA5 Reanalysis data from the European Center for Medium-range Weather Forecasts (ECMWF) (Hersbach et al., 2020); (2) Calendrical context data, relative to the specific type of day of the prediction; (3) Persistence related data, including information on users from previous days or periods of time; and finally (4) analogous instances data, as previously mentioned, it consists of including as predictor variables those with information about analogous instances to the current prediction.

The specific predictor variables used for each problem can be consulted in Sections 2.1–2.3. Also, an analysis of the demand profile evolution for the different databases as a function of the predictive features provided in Section 2.4.

2.1. BiciMad

BiciMad[1] is a product by Avanza Bike S.L., a private corporation stationed in Madrid, for e-bicycle rentals. According to their official website, corporation offers almost three thousand e-bicycles and more than six thousand parking lots. Since the beginning of the business, BiciMad is collecting the data on a daily basis, consisting of the total number of bikes rented per day. This dataset is public, and can be found and downloaded from Ayuntamiento de Madrid (2022a).

The database has been extended with 19 predictor variables listed in Table 1 along with their descriptive statistics. These variables consist of meteorological information for the selected time period, information about preceding days and weeks, together with information about analogous days to the current one.

[1] https://www.bicimad.com/.

2.2. Madrid Cable Car

Madrid Cable Car (*Teleférico de Madrid*) belongs to the public transport company of Madrid (EMT, i.e. Empresa Municipal de Transportes de Madrid, S. A.). Specifically, Teleférico is a provider of a cable car transport on the west of the Madrid, currently holding 80 cabins, each of them able to carry 6 people. The goal of the dataset is to predict number of total users per day. Data of Madrid cable car demand are also public, and can be found and downloaded in Ayuntamiento de Madrid (2022b).

As previously mentioned, this dataset contains dispersed data with several days with missing data. As the percentage of instances with no data is significant (40%), it was decided not to fill these days with dummy values, but rather to deal with the incomplete dataset. Therefore, instead of using as predictor variables those related to users of previous days or weeks, which are not available for all instances, the weekly and monthly averages were used. Hence, the total number of predictor variables is lower than for the other two problems, with just 15 features, whose statistical information can be consulted in Table 2.

2.3. Bicing Barcelona

Bicing Barcelona is a public rental service in the city of Barcelona, Spain, which was implemented in March 2007, promoted by the City Council. The city company *Barcelona de Serveis Municipals (BSM)* is the manager of the service, while the operating concession corresponds since 2019 to *Pedalem Barcelona*. Between 2007 and 2018, Clear Channel was the company in charge of this service. Currently it includes a total of 7000 bicycles and an average daily use of 5.54 uses per bicycle and day. The dataset contains information about the number of bicycles in use during each hour of the day by intervals of 15 min from August 2018 to January 2019. Data of bikes demand in Barcelona are public as the previous datasets considered, and can be found and downloaded in Ayuntamiento de Barcelona (2022).

The regression problem consists in this case of predicting the average number of bicycles in use per hour. The predictor variables include a total of 29 features, providing information on both previous and analogous days, and previous and analogous hours. The descriptive statistics of this variables can be consulted in Table 3.

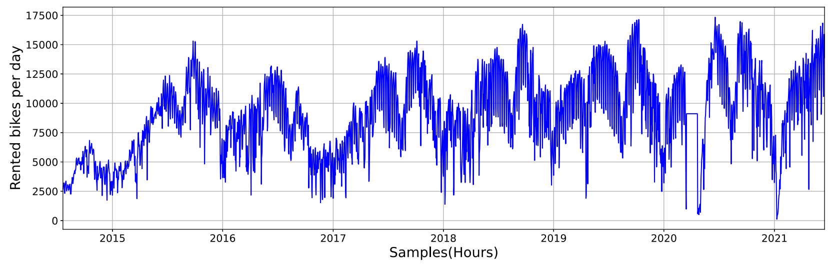
2.4. Mobility demand profiles

As part of a preliminary analysis of the raw data, the demand profiles for the three cases of urban mobility are shown in Fig. 2, as a function of the different forecasting variables. In order to allow a comparison between the different databases under study, the demand values have been normalized between 0 and 1 for each case.

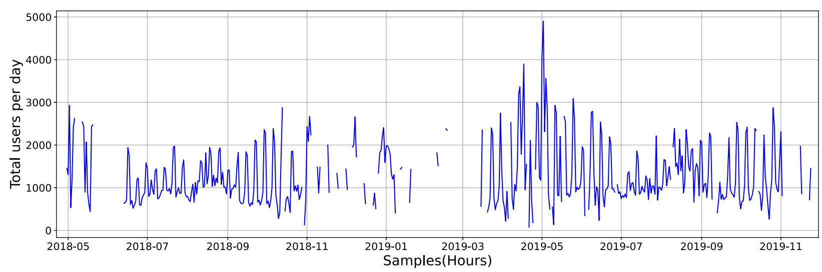
Here, it is possible to observe certain significant disparities in the trend of the two transportation means assessed, i.e.: bike sharing and cable car. While the bicycle use decreases on weekends, the number of cable car passengers increases substantially on those days. The same trend is observed during festivities, where both forms of transport exhibit opposing profiles. This leads us to conclude that the use of shared bicycles corresponds to a work-related profile, while the cable car may be considered a leisure activity.

Furthermore, the figure regarding the demand profile according to the month of the year reveals that the use of bicycles is higher in the months with better weather: between May and October (with the exception of August, when an important urban center as Madrid gets emptier due to the holiday period). Concerning the cable car, however, an opposite trend is observed, being precisely in the warmer months (from June to October) when the demand is lower. This graph is only shown for BiciMad and Madrid Cable Car dataset, due to the fact that the Bicing Barcelona database only has a duration of 3 months.

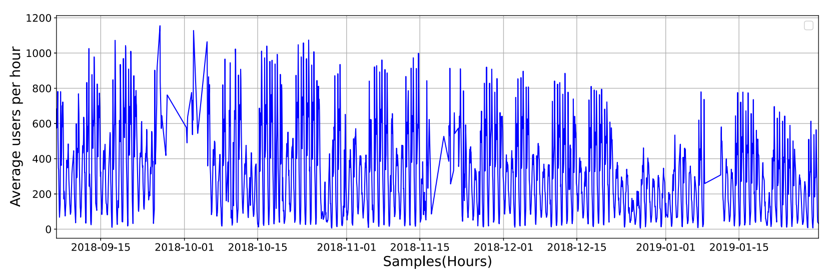
Regarding the demand variation according to the time of day, it can be seen in Bicing Barcelona (the only database with hourly frequency)



(a) BiciMad time series.



(b) Madrid's cable car time series.



(c) Bicing Barcelona time series.

Fig. 1. Timed series considered in this work.

that the demand is higher in the early morning (from 7 am to 9 am) and in the afternoon (from 2 pm to 8 pm), confirming that it is a transportation mode usually utilized by local people to travel to their respective places of work/study.

Finally, the analysis of figures showing the demand profiles as a function of meteorological variables reveal, firstly, an increase in demand as temperature rises in the case of bike-sharing services. For the cable car, nonetheless, it is observed that with very high temperature values, demand decreases again. Secondly, both databases share the fact that a decrease in demand occurs as precipitation level increase, as somehow expected with bad weather.

3. Machine learning and deep learning regression methods

In this section, a short description of ML/DL methods employed in the experiments section is provided. Note that we do not intend a comprehensive description of the ML/DL methods considered, but a brief description with enough references, so the interested reader can consult details on them.

3.1. Regression models

Regression is a popular, yet traditional, modeling method, often exploited as a benchmark method. The linear regression (LR) originates from the famous mathematician (Gauss, 1823), who implemented the straightforward and deterministic ordinary least squares (OLS) method for modeling problems. Plenty of alternatives emerged since then, among them also several state-of-the-art regression methods, such as polynomial regression (PR) (Ostertagová, 2012) and lasso regression (Tibshirani, 1996) (Lasso). Compared to the linear regression, polynomial regression can capture higher (nth) orders of relationships between explanatory variable(s) and a dependent variable, by transforming the explanatory variables non-linearly. Still, an OLS method can be exploited to fit the relationship afterwards. Lasso on the other hand performs so-called L1 regularization, i.e., shrinking calculated regression coefficients towards zero by penalizing them using a combination of residual sum of squares and the shrinkage penalty term. The goal of the shrinking regression coefficients is to obtain a reduced variance of coefficients, where shrinking is conveniently controlled by a shrinkage parameter (usually denoted by λ; original Tibshirani's equation Reid, Tibshirani, & Friedman, 2016 specified as $1/2\|Y - X\beta\|_2^2 + \lambda\|\beta\|_1$, where the $Y - X\beta$ represents the residual sum of squares). In specific cases, lasso regression may act as a Feature Selection (FS) procedure, which makes it especially useful for highly-correlated data, and may help at preventing overtraining (overfitting).

3.2. Classification and Regression Trees (CART)

Classification and Regression Trees are an umbrella term for both the (1) classification trees, i.e. those that perform classification tasks,

Table 1
Descriptive statistics of the BiciMad dataset.

Var.	min	max	stdev	kurt	skew	type
DayofWeek	1	7	2.000	−1.250	0.001	integer
Holiday	0	1	0.193	20.981	4.792	dummy
Month	1	12	3.479	−1.226	−0.036	integer
PreviousDayBikes	118	17 338	3478.326	−0.675	−0.030	integer
PreviousWeekAverage	736	16 001	3030.740	−0.630	−0.233	float
PreviousMonthAverage	1220	14 975	2854.314	−0.632	−0.258	float
SameDayPreviousWeek	118	17 338	3479.183	−0.674	−0.031	integer
SameDay4WeeksAverage	969	16 924	3169.546	−0.637	0.023	float
Temperature	276	312	8.337	−1.102	0.224	float
Rainfall	0	4734	240.093	138.079	9.820	float
PreviousDayTemperature	276	312	8.336	−1.101	0.225	float
PreviousDayRainfall	0	4734	240.150	137.930	9.812	float
NextDayHoliday	0	1	0.193	20.981	4.792	dummy
NextDay_DayofWeek	0	7	2.002	−1.249	−0.001	integer
LastAnalogDay	99	16 873	2966.583	−0.389	0.006	float
AverageAnalogDays	99	17 146	3515.943	−0.707	−0.009	integer
2DaysAgo	118	17 338	3475.810	−0.674	−0.031	integer
3DaysAgo	118	17 338	3476.542	−0.673	−0.032	integer
4DaysAgo	118	17 338	3478.784	−0.675	−0.031	integer
RentedBikes	118	17 338	3480.853	−0.674	−0.033	integer

Note: Data is available from 21st August 2014–16th June 2021, with a sample size $N = 2486$. DayOfWeek: 1 = Monday, 7 = Sunday, Holiday: today, 0 = No holiday, 1 = holiday, Month: 1 = January, 12 = December, PreviousDayBikes: no. of bike rentals yesterday, PreviousWeekAverage: no. of bike rentals in a previous week in average, PreviousMonthAverage: no. of bike rentals in a previous month in average, SameDayPreviousWeek: no. of bike rentals in previous week on the same day of week, SameDay4WeeksAverage: no. of bike rentals on 4 previous consecutive weeks on the same day of week in average, Temperature: daily average temperature, °K, Rainfall: mm/24 h, PreviousDayTemperature: daily average temperature, yesterday °K, PreviousDayRainfall: previous day rainfall, mm/24 h, NextDayHoliday: tomorrow, 0 = No holiday, 1 = holiday, NextDay_DayofWeek: tomorrow 0 = last day of the dataset, 1 = Monday, 7 = Sunday, LastAnalogDay = no. of bike rentals on last similar day, AverageAnalogDays = no. of bike rentals on analogous days on average, 2DaysAgo: no. of bike rentals 2 days ago, 3DaysAgo: no. of bike rentals 3 days ago, 4DaysAgo: no. of bike rentals 4 days ago, RentedBikes: no. of bike rentals in a given day, dependent variable.

Table 2
Descriptive statistics of a Madrid cable car dataset.

Var.	min	max	stdev	kurt	skew	type
DayofWeek	1	7	2.044	−1.291	−0.178	integer
Holiday	0	1	0.218	15.358	4.157	dummy
Temperature	277.401	312.346	7.441	−0.667	−0.522	float
Rainfall	0	4733.563	266.751	242.410	14.365	float
Month	1	12	2.477	−0.193	−0.383	integer
PreviousMonthAverage	874	2102	271.491	0.610	1.088	float
PreviousWeekAverage	563	2882.286	382.647	3.314	1.539	float
PreviousDayTemperature	277.401	312.346	7.485	−0.671	−0.501	float
PreviousDayRainfall	0	4733.563	274.604	217.267	13.511	float
Rainfall_asCat	0	1	0.437	−0.747	1.121	dummy
PreviousDayRainfall_asCat	0	1	0.443	−0.876	1.062	dummy
NextdayHoliday	0	1	0.201	18.941	4.566	dummy
NextDay_DayofWeek	0	7	2.130	−1.362	−0.027	integer
LastAnalogDay	0	3898	554.292	2.282	1.415	float
AverageAnalogDays	0	2996	528.291	1.131	1.206	float
TotalUsers	75	4902	718.667	1.744	1.201	integer

Note: Data is available for period 22nd April 2018–30th November 2019, sample size $N = 402$. DayOfWeek: 1 = Monday, 7 = Sunday, Holiday: today, 0 = No holiday, 1 = holiday, Temperature: daily average temperature °K, Rainfall: mm/24 h, Month: 1 = January, 12 = December, PreviousMonthAverage: no. of users in a previous month in average, PreviousWeekAverage: no. of users in a previous week in average, PreviousDayTemperature: daily average temperature, yesterday°K, PreviousDayRainfall: previous day rainfall, mm/24 h, Rainfall_asCat: rainfall today in a dummy form, 0 = no rainfall, 1 = rainfall, PreviousDayRainfall_asCat: rainfall yesterday in a dummy form, 0 = no rainfall, 1 = rainfall, NextDayHoliday: tomorrow, 0 = No holiday, 1 = holiday, NextDay_DayofWeek: tomorrow 0 = last day of the dataset, 1 = Monday, 7 = Sunday, LastAnalogDay = users on last similar day, AverageAnalogDays = users on analogous days on average, TotalUsers: no. of users in a given day, dependent variable.

and (2) regression trees (RT), i.e. those that perform regression tasks. CART methods have in past sustained a dynamic and vigorous research forward, as were a subject of many modifications and improvements. In general, CARTs are hierarchical splitting decision-rule based systems, where their complexity varies with the depth of the trees (the deeper the trees the more detailed the modeling). Traditionally, an impurity measure, e.g., Gini impurity, is calculated as a trial-and-error learning signal. General CART methods are prone input variation stability, hence these can be upgraded to include several instances of trees. Such example include for instance ensemble-based methods, such as Random Forests (RF) (Breiman, 2001; Ho, 1995) (Fig. 3). Here, several trees are fitted independently using a random sampling with replacement and the voting mechanism is employed upon to deliver the most represented result. Indeed, the more the trees in a forest, the longer the time complexity of training (fitting) and prediction.

3.3. Support Vector Machine (SVM)

SVMs are a state-of-the-art classification and regression modeling and prediction methods. The core of the SVMs are the so-called kernel functions that can be either linear or non-linear. A modern version of SVM, as proposed in 1995 by Cortes and Vapnik (Cortes & Vapnik, 1995), are in general divided into two common subgroups, (1) the SVC or support vector classification and (2) the SVR or support vector

Table 3
Descriptive statistics of the Bicing Barcelona dataset.

Var.	min	max	stdev	kurt	skew	type
DayofWeek	0	6	2.032	−1.295	0.002	integer
Holiday	0	1	0.249	10.165	3.488	dummy
Hour	0	23	6.811	−1.156	−0.089	integer
Month	1	12	3.839	0.462	−1.443	integer
PreviousDaySameHourAverage	0	1073.750	254.606	−0.218	0.793	float
PreviousWeekSameHourAverage	0	1109.500	207.995	−0.734	0.263	float
PreviousMonthSameHourAverage	32.302	821.633	193.338	−0.961	0.074	float
SameDaySameHourPreviousWeek	0	1155	260.341	−0.259	0.816	float
SameDaySameHour4WeeksAverage	142.614	545.352	94.732	−0.949	−0.007	float
PreviousDayAverage	0	801.750	124.284	−0.335	0.162	float
PreviousWeekAverage	166.236	666.663	81.004	1.608	0.146	float
PreviousMonthAverage	223.018	432.845	55.328	−0.524	−0.436	float
SameDayPreviousWeek	0	801.750	132.254	0.352	0.274	float
SameDay4WeeksAverage	166.803	524.105	89.461	−1.010	−0.030	float
PreviousHourAverage	0	1155	253.632	−0.250	0.759	float
Previous2HourAverage	0	1155	234.414	−0.445	0.581	float
Previous4HourAverage	0	1155	211.301	−0.622	0.428	float
Previous12HourAverage	0	1155	157.507	0.001	0.541	float
Temperature	274.864	301.883	5.768	−0.715	0.387	float
PreviousDaySameHourTemperature	274.864	301.792	20.781	−0.345	−1.283	float
PreviousDayAverageTemperature	274.864	298.230	20.105	−0.184	−1.315	float
SameDayAverageTemperature	278.962	298.230	5.264	−0.902	0.590	float
NextDayAverageTemperature	278.962	298.230	18.307	2.200	−14.254	float
LastAnalogHour	6.417	1155	238.254	−0.208	0.663	float
AverageAnalogHours	8.500	1018.528	219.312	−0.402	0.485	float
LastAnalogDay	6	1155	236.234	−0.152	0.686	float
AverageAnalogDays	36.271	559.250	83.496	−0.413	−0.629	float
NextDayHoliday	0	1	0.245	10.653	3.557	dummy
NextDay_DayofWeek	0	6	2.028	−1.291	0.044	integer
AverageRentedBikes	6	1155	236.349	−0.151	0.690	integer

Note: Data is available for period 6th September 2018–30th January 2019, sample size $N = 2400$. DayOfWeek: 0 = Monday, 6 = Sunday, Holiday: today, 0 = No holiday, 1 = holiday, Hour: 0 = 0 am, 23 = 11 pm, Month: 1 = January, 12 = December, PreviousDaySameHourAverage: Av. bike rentals in the previous day at same hour, PreviousWeekSameHourAverage: Av. bike rentals in the previous week at same hour, PreviousMonthSameHourAverage: Av. bike rentals in the previous month at same hour, SameDaySameHourPreviousWeek: Av. bike rentals in the same day of week at previous week and same hour, SameDaySameHour4WeeksAverage: Av. bike rentals on 4 previous consecutive weeks on the sameday of week and same hour, PreviousDayAverage: Av. bike rentals in the previous day, PreviousWeekAverage: Av. bike rentals in the previous week, PreviousMonthAverage: Av. bike rentals in the previous month, SameDayPreviousWeek: Av. bike rentals in previous week on the same day of week, SameDay4WeeksAverage: Av. bike rentals on 4 previous consecutive weeks on the sameday of week, PreviousHourAverage: Av. bike rentals in the previous hour, Previous2HourAverage: Av. bike rentals in the previous 2 h, Previous4HourAverage: Av. bike rentals in the previous 4 h, Previous12HourAverage: Av. bike rentals in the previous 12 h, Temperature: hourly temperature, K, PreviousDaySameHourTemperature: Previous day temperature at same hour, PreviousDayAverageTemperature: Previous day av temperature, SameDayAverageTemperature: Current day av temperature, NextDayAverageTemperature: Next day av temperature, LastAnalogHour: Av. bike rentals on last similar hour, AverageAnalogHours: Av. bike rentals on analogous hours, LastAnalogDay: Av. bike rentals on last similar day, AverageAnalogDays: Av. bike rentals on analogous days, NextDayHoliday: tomorrow, 0 = No holiday, 1 = holiday, NextDay_DayofWeek: 0 = Monday, 6 = Sunday, AverageRentedBikes: Av. bike rentals in a given day, dependent variable.

regression (Smola & Schölkopf, 2004). The philosophy of the SVM method lies in finding for the so-called support vectors and placing the so-called separating hyperplane(s) between them, dividing the input datapoints into one or more classes. Naturally, datapoints treated as outliers are present in majority of practical cases and these may impose significant (misleading) effects toward positioning hyperplanes. However, a so-called soft-margin be employed to allow blended positioning of support vectors near separating hyperplanes, increasing SVM reliability. Noble (2006) stated that any consistent data may be separated using a linear hyperplane if projected into higher-enough dimensional space, but due to the curse of dimensionality, solving such problems increases complexity drastically. Hence, a non-linear kernel functions may be proposed. Still, no free lunch theorem prevails here, leaving the practical evaluation of a set of kernel functions unavoidable. Due to its straightforward applicability and suitability, SVM has in past been tailored to a domain of financial prediction (Trafalis & Ince, 2000), text classification (Tong & Koller, 2001), medicine (Huang et al., 2018), and many others.

3.4. Artificial Neural Networks (ANNs)

Multilayer Perceptron (MLP) (Gardner & Dorling, 1998) are simple, yet robust, connectionist systems, that incorporate a layered structure of artificial neurons (perceptrons) and artificial synapses, i.e. connections between perceptrons. Typically, MLP consists of an input layer, at least a single (or more) hidden layers, and an output layer (Fig. 4). MLP architectures can be configured to deliver either regression or classification tasks. MLPs are stochastic methods, a backpropagation trial-and-error is among the more well-known learning algorithms applied to train ANNs.

Traditionally, ANNs are learnt (tuned) in an iterative way using the either gradient or momentum-based learning algorithms, which in some cases require massive processing and memory capabilities. As a result, an Extreme Learning Machine (ELM), i.e. type of the ANNs with the specialized learning algorithm that are tuned in a single step (Huang, Zhu, & Siew, 2006). ELMs are a revolutionary way of ANN learning and has since inception attracted many researchers due to its straightforwardness and efficient time complexity. A survey of application (Ding, Xu, & Nie, 2014) shows that ELMs have been successfully applied to a sample of tasks, such as classification/regression, pattern recognition, forecasting/diagnosing, image processing, among many others (see Fig. 5).

3.5. Deep learning methods

Recurrent Neural Networks (RNNs) are a specialized type of ANNs that incorporate a feedback loop, i.e. the output of the RNN layer

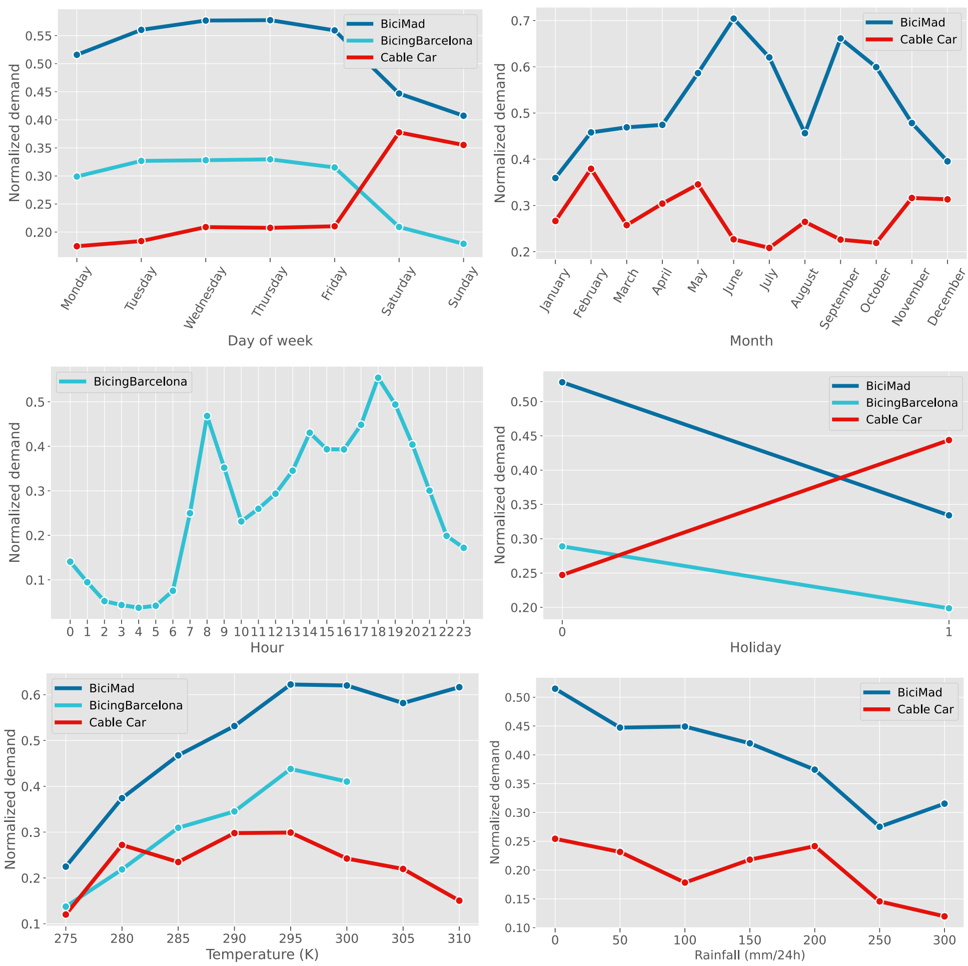


Fig. 2. Mobility demand profiles according to different predictive features.

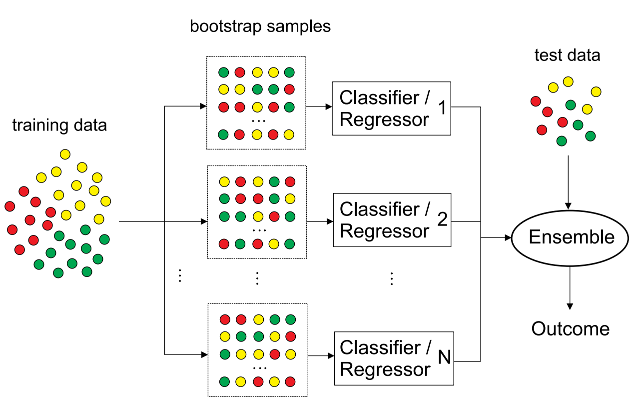


Fig. 3. Bagging technique for classification or regression problems.

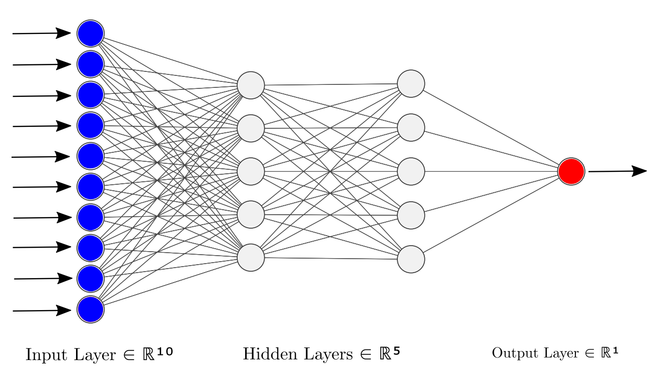


Fig. 4. Structure of a MLP, with two hidden layers.

(again) taken as an input (Fig. 6). Hence, RNNs are especially viable for analyzing time-series and sequence data. Traditionally, a truncated backpropagation through time (BPTT) is exploited for fitting the data (Lillicrap & Santoro, 2019). RNN networks were proposed by Rumelhart, Hinton, and Williams (1986) and has since its inception been successfully applied in various areas, such as text generation (Sutskever, Martens, & Hinton, 2011), speech recognition (Graves, Mohamed, & Hinton, 2013), sequence generation (Graves, 2013), among others.

Inspired by Hochreiter and Schmidhuber (1997), Long Short-Term Memory (LSTM) networks are a type of an RNN network, with additional integrated feature, i.e. the memory cell. The memory cell, through the implementation of input, output and forget gates, efficiently deals with the problem of vanishing gradient (that RNNs are prone to) and thus allows the LSTMs to capture learnt knowledge for a much longer time (Fig. 7). A dedicated forgetting and filtering methods are employed here prior to writing new information into the memory cell to allow selecting most distinctive data that maximizes the learning and memory capabilities. LSTMs are specifically designed to cope with the sequence data, hence some modifications need to be made prior using the original dataset. Typical applications of LSTMs

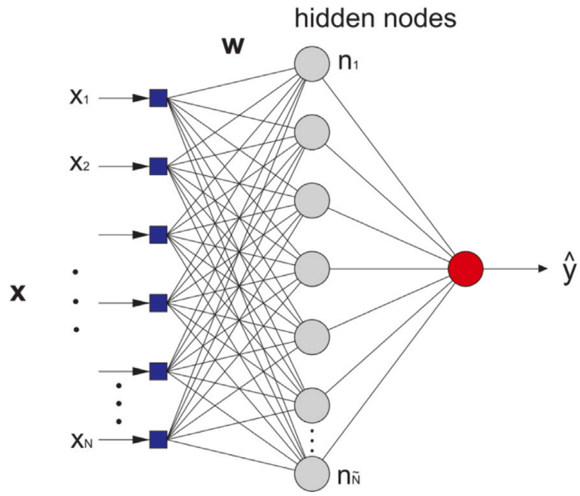


Fig. 5. Structure of an ELM network.

can be found in sentence embedding (Huang et al., 2020; Palangi et al., 2016), machine reading (Cheng, Dong, & Lapata, 2016), automated stock trading (Fister, Perc, & Jagrič, 2021), sentiment analysis (Islam, Datta, & Iqbal, 2023), energy consumption forecasting (Peng, Wang, Xia, & Gao, 2022), meteorological events forecasting (Peláez-Rodríguez et al., 2023), etc.

Gated Recurrent Units (GRU) networks (Chung, Gulcehre, Cho, & Bengio, 2014) emerged in 2014 as a streamlined version of the LSTM, with a simplified memory cell that considerably reduces the high computational cost of the LSTM and achieves comparable performance. GRU layers only present two gates: (1) the update gate, which decides whether the memory state is or is not updated; and, (2) the relevance gate, which determines how relevant is to compute the next candidate (Fig. 8).

Convolutional Neural Networks (CNN) (Fukushima & Miyake, 1982) are a specific type of feed-forward neural networks initially developed for tasks related to image processing and computer vision. CNN architecture typically consists of a sequential stacking of layers (Krizhevsky, Sutskever, & Hinton, 2017). Convolution layers are responsible for the learning of the features from input data. They apply and slide a filter over the data. This filter, also known as kernel, contains learnable weights and biases, and is the equivalent of nodes in a regular neural network layer. Once the features have been extracted by the convolutional layers, the forecasting is carried out by using fully connected neural network. The input data for these fully connected layers are the flattened features resulting of the convolution (Fig. 9). In recent years, CNNs have become increasingly popular not only for image processing tasks, but also for all kinds of sequential data related problems, such as text or time series forecasting. Usually, one-dimensional CNN (CNN1D) are applied to extract features for each variable time-series (Ismail Fawaz, Forestier, Weber, Idoumghar, & Muller, 2019).

4. Experiments and results

After preparing and preprocessing the datasets, the pool of 12 ML/DP methods was evaluated for each of the 3 regression problems. In addition, a sensitivity analysis has been performed on each problem, in order to determine the importance of each group of variables in the regression. Finally, a feature selection algorithm was applied and its influence on the results was evaluated. The quality of predictions was assessed using four statistical regression metrics presented in detail in Section 4.1. Results are summarized in tabular and graphical forms.

4.1. Regression metrics

Four commonly used regression metrics have been employed to assess the performance of the ML/DL methods applied to the three proposed problems: the Pearson correlation coefficient (R^2, also called coefficient of determination), the Root Mean Squared Error (RMSE), the Mean Absolute Percentage Error (MAPE) and the Mean Absolute Error (MAE). The correspondent equation of each metric is shown in Eq. (1), where $\hat{y}$ represents predicted values (provided by the model) and y are the actual values. The subscript i is used to refer to a single sample $y_i = y[i]$.

$$R^2 = \frac{\sum_{i=1}^{n}(y_i - E[y])(\hat{y}_i - E[\hat{y}])}{\sqrt{\sum_{i=1}^{n}(y_i - E[y])^2}\sqrt{\sum_{i=1}^{n}(\hat{y}_i - E[\hat{y}])^2}} \tag{1a}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{1b}$$

$$\text{MAPE} = \frac{100}{N}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{1c}$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{1d}$$

Note that the higher the R^2, the greater the goodness-of-fit and the better the predictions. On the other hand, the lower the RMSE, MAPE, MAE, the lower the prediction error and the better the predictions obtained.

4.2. Data preprocessing

Data preprocessing step consisted of four consecutive sub-steps: (1) adapting the target variable, as the original (raw) data in each dataset incorporated some out-of-scope characteristics, such as types of users or bicycles. Instead, these were transformed into the total number of users per period of time; (2) scoping and analyzing the explanatory variables, these can be classified into four groups attending to the type of information of each variable (meteorological, categorical, previous day and analog instances, Fig. 10). Also, Figs. 11–13 show the relative correlation coefficients among each group of variables, where similar patterns are found for the 3 databases: a very high correlation is observed among the variables belonging to the groups of previous day and analog instances variables, respectively. In the case of the meteorological variables, there is a strong correlation between the variables related to temperature and between those related to precipitation, but they are not mutually correlated. Finally, no strong correlation is observed among the categorical variables for any database.; (3) scaling the features, which is important to ensure the upper and lower limits of data in the given predefined range. Feature standardization was performed, causing data to have zero-mean and a unit-variance (Eq. (2)), as follows:

$$x' = \frac{x - \bar{x}}{\sigma} \tag{2}$$

where x is the original feature vector, $\bar{x}$ denotes the feature mean and σ its standard deviation. Finally (4), a training-test split (80%–20%), considering the in-sample and out-of-sample was executed, assuring that no test instance was seen by the ML/DL method during the training. Since dealing with timed-series data, instead of randomly splitting the datasets, last 20% of instances have been removed to validate the methods.

4.3. Feature selection

The number of predictor variables is something of vital importance for an efficient performance of the regression methods, otherwise introducing too many closely related variables may cause problems of redundancy or high level of multi-collinearity.

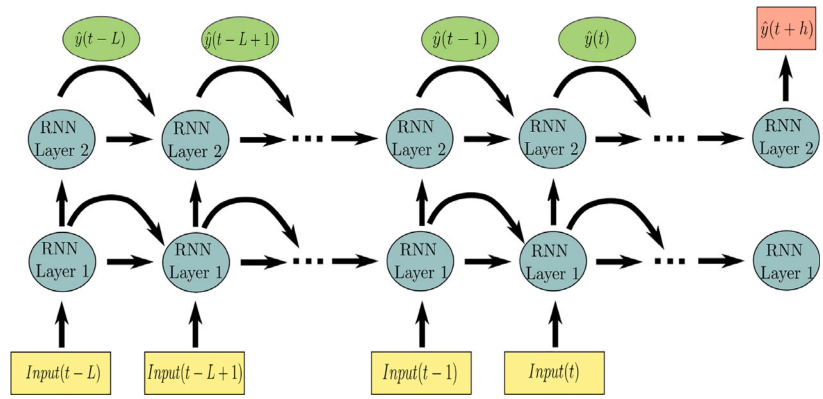


Fig. 6. Architecture of a deep RNN for a number of RNN layers equal to 2.

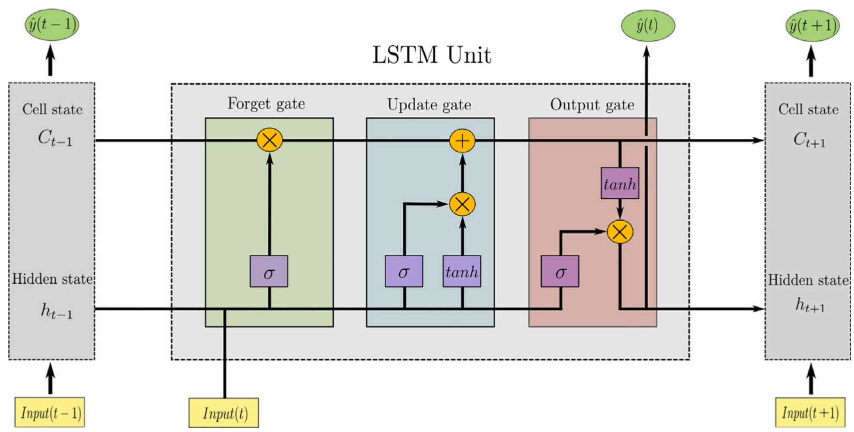


Fig. 7. Architecture of a LSTM network.

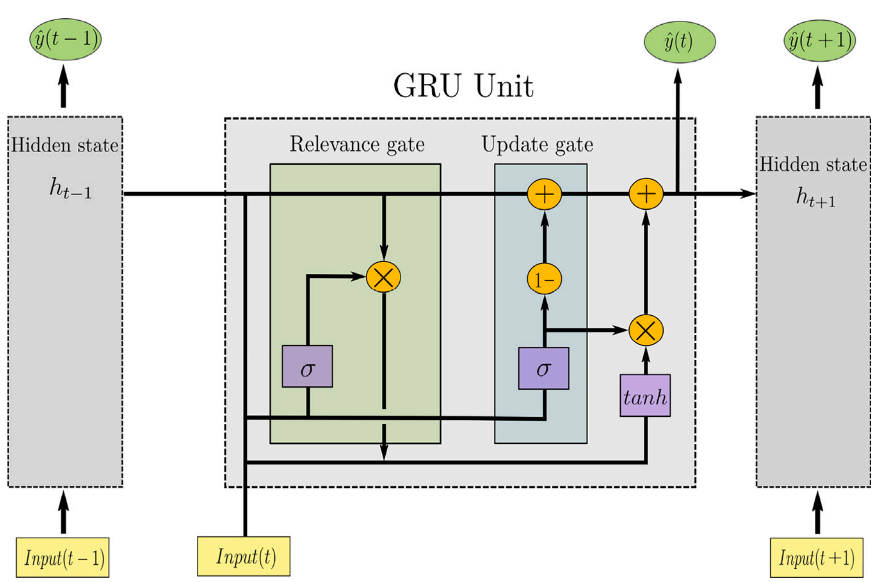


Fig. 8. Architecture of a GRU network.

Besides, it was observed that each of the implemented methods had a different set of optimal predictor variables. Therefore, instead of relying on a complex FS method that selects a single set of variables for all the methods, a simple exhaustive search algorithm was applied, associating with each regression technique the set of input variables with the lowest error in the training set. It was also observed during a preliminary analysis that some methods performed better when entering the labels of day of the week, month or hour as numerical variables (integers), while other methods preferred to work with categorical dummy variables. Hence, in the implemented FS algorithm, the optimal representation of these kind of variables is also sought for each method.

Nevertheless, as the number of predictor variables is relatively high (19, 15 and 29), it is not convenient to perform the exhaustive search

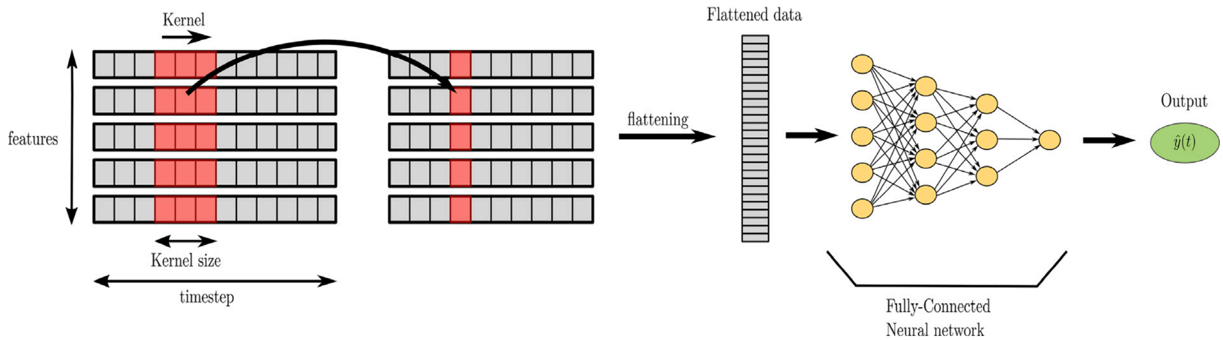


Fig. 9. Architecture of a 1D-CNN.

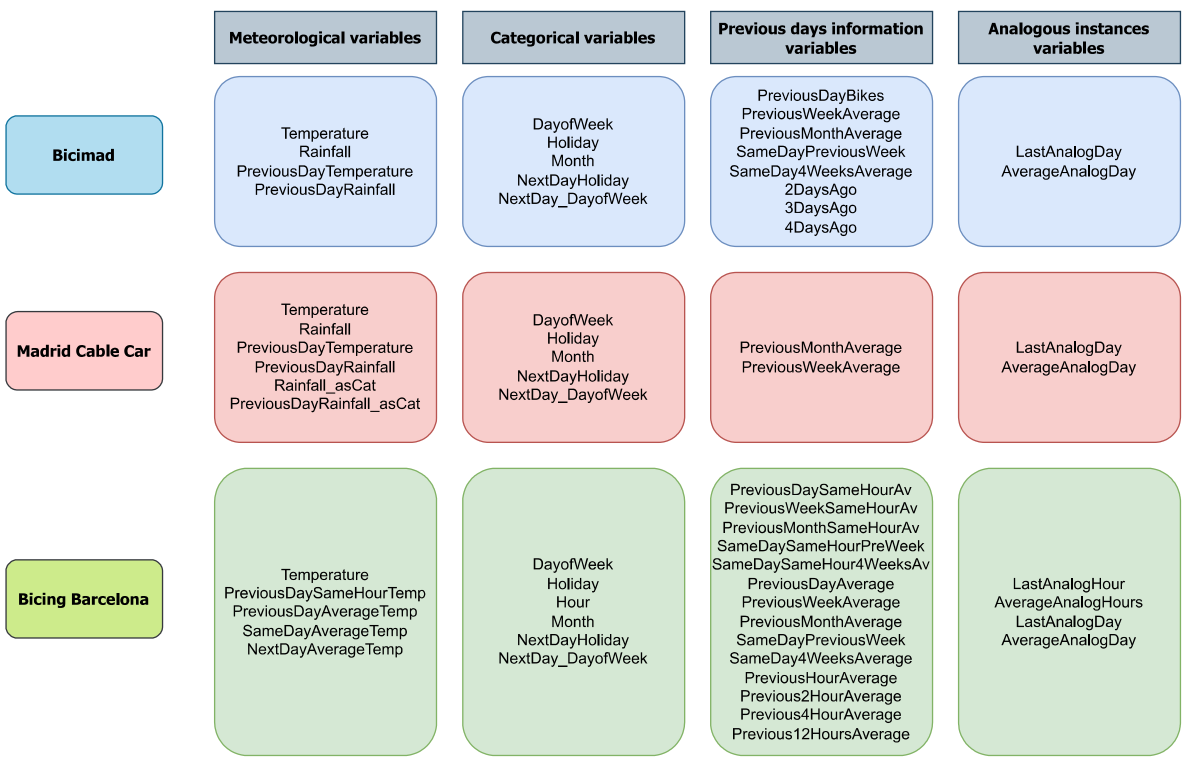


Fig. 10. Predictor variables for the three datasets divided by type of information.

process with all the predictor variables, since the number of possible combinations would amount 2^{19+2}, 2^{15+2} and 2^{29+3}, respectively, and this procedure would need to be repeated for each of the 12 regression methods applied. This makes the application of a pure brute force algorithm unsuitable, considering that some implemented techniques (MLP, SVR or deep methods) presents training times considerably long.

Instead, it was preferred to retain a certain number of variables as fixed, those considered of great relevance for the prediction outcome, and to perform the exhaustive search procedure with the remaining non-fixed features. In this way, both the computational effort and the time required for the execution of the algorithm are significantly lightened. The selection criteria for choosing the set of fixed features in each problem consisted of calculating the Pearson correlation coefficient between the target variable and the predictors for each dataset, and setting aside 60% of variables with the highest correlations in absolute terms as fixed variables. The correlation coefficients for each of the regression problems can be consulted in Tables 4–6. A preliminary analysis of these tables shows the high correlation between the target variable and the predictor variables related to both analogous instances and information related to previous days in the three datasets, so one would expect these features to be of great importance for the regression methods. It is noteworthy to observe how meteorological variables affect the number of users of each problem, since a significant relationship is appreciated between temperature and the number of bicycles rented in both Madrid and Barcelona, with a correlation of 0.4 in both cases. In the Cable Car, however, temperature does not seem to have the same relationship with the number of users, with a negative correlation coefficient. Regarding the influence of rainfall, a negative correlation is observed in the two problems in which it has been included as a predictor variable, this is something to be expected, since it seems logical that the use of outdoor urban transport decreases on rainy days.

Thus, the number of fixed variables in each of the problems has been set at 11, 9 and 17, respectively, so that the number of combinations in the exhaustive search amounts to 2^{8+2}, 2^{6+2} and 2^{12+3}, which is considerably lower than previously reported.

The exhaustive FS procedure was formalized as follows: a boolean attendance vector $\mathbf{a}_i$ of dimension $n + m$, where n represents total number of features and m denotes the number of features to be tested both as numerical and as categorical variables, was initialized with the first j elements as ones and the remaining $k + m$ elements as zeros, where j represents the set of features considered fixed for each problem

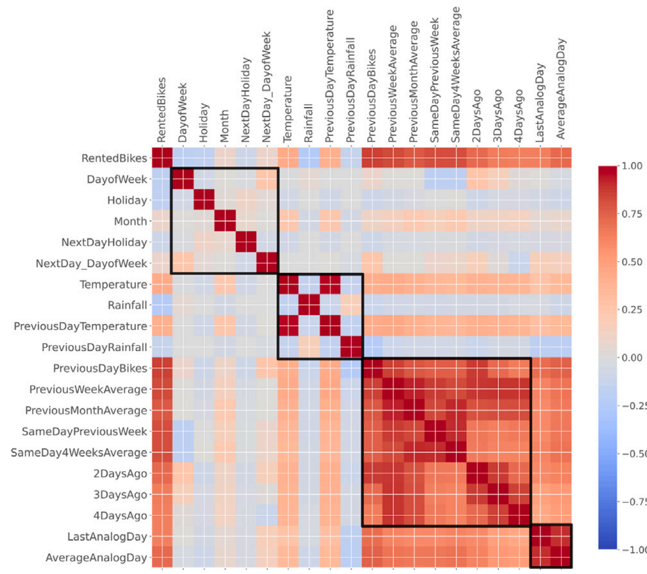


Fig. 11. Correlation coefficients among the variables belonging to the BiciMad dataset, where black squares corresponds to each of the four groups of predictor variables.

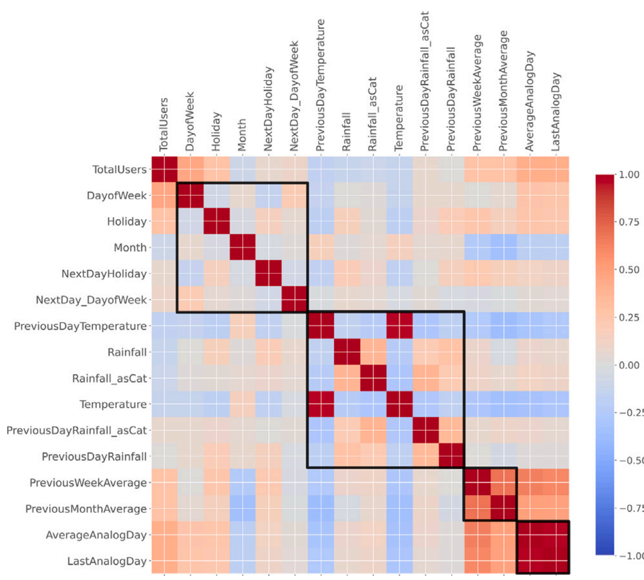


Fig. 12. Correlation coefficients among the variables belonging to the Madrid's Cable Car dataset, where black squares corresponds to each of the four groups of predictor variables.

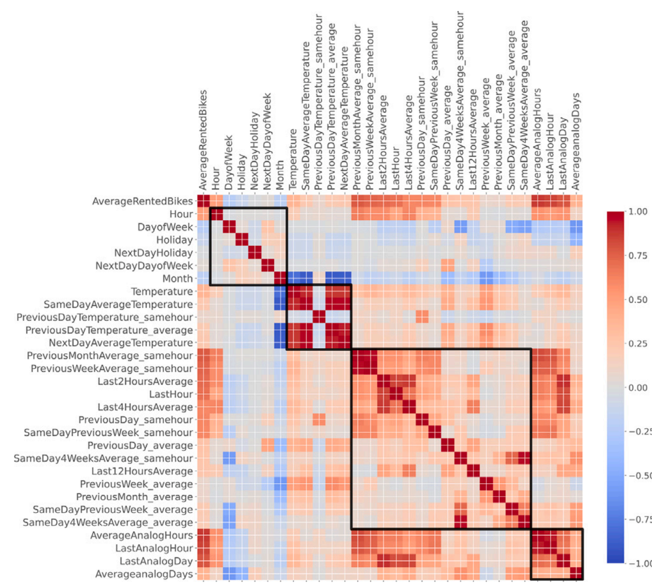


Fig. 13. Correlation coefficients among the variables belonging to the BicingBarcelona dataset, where black squares corresponds to each of the four groups of predictor variables.

and k denotes the remaining non-fixed variables. Then, a derivation of all possible (boolean) combinations of $k+m$ was executed. At each iteration, the corresponding regression model is trained with the entry variables defined by the vector $\mathbf{a}_i$, where for the first n elements one denotes the presence and zero denotes the absence of the specific feature, and for the last m elements one denotes that the feature is represented as a categorical dummy variables and zero indicates that is represented as a numerical one. Once the feature set for a given iteration has been defined, the corresponding ML model is trained with the training data updated to the selection of features defined in that iteration. The training error coefficients are computed for the specific model, and compared to the best error achieved so far. The algorithm is programmed to minimize the training MAE, so if the MAE coefficient of the model at iteration i is lower than the minimum MAE achieved, both the best_model and the best_MAE variables are updated. A descriptive diagram flow of the FS algorithm presented is shown in Fig. 14.

Table 4
Correlation coefficients between the target and predictor variables for the BiciMad dataset, sorted in descending correlation order.

Predictor variables	R	Predictor variables	R
PreviousDayBikes	0.8508	Temperature	0.4195
PreviousWeekAverage	0.8203	PreviousDayTemperature	0.4000
SameDayPreviousWeek	0.8031	Month	0.1138
SameDay4WeeksAverage	0.7723	NextDay_DayofWeek	0.0771
PreviousMonthAverage	0.7391	NextDayHoliday	−0.0969
2DaysAgo	0.7027	PreviousDayRainfall	−0.1251
AverageAnalogDays	0.7013	Holiday	−0.1831
LastAnalogDay	0.6353	DayofWeek	−0.2000
3DaysAgo	0.6333	Rainfall	−0.2006
4DaysAgo	0.6118		

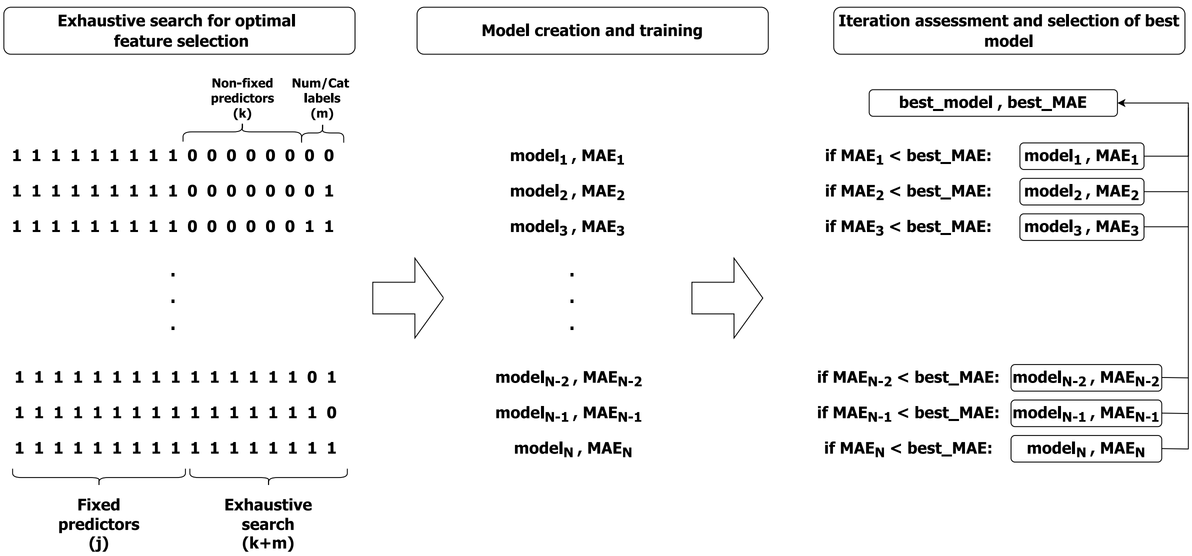


Fig. 14. Implemented algorithm for selection of optimal predictor variables for each ML technique.

Table 5
Correlation coefficients between the target and predictor variables for the Madrid's Cable Car dataset, sorted in descending correlation order.

Predictor variables	R	Predictor variables	R
DayofWeek	0.4783	PreviousDayRainfall_asCat	0.0298
AverageAnalogDays	0.4355	PreviousDayRainfall	−0.0578
LastAnalogDay	0.4249	Month	−0.0909
Holiday	0.2875	Temperature	−0.1136
PreviousWeekAverage	0.2507	Rainfall_asCat	−0.1162
PreviousMonthAverage	0.2446	Rainfall	−0.1233
NextDay_DayofWeek	0.0878	PreviousDayTemperature	−0.1492
NextdayHoliday	0.0533		

4.4. Experimental setup and hyperparameter tuning

The experimental setup is shown in Table 7. Hyperparameters for ML methods were not set, but instead a grid- or randomized-search hyperparameter tuning has been employed. The former was used in case that two hyperparameters were tuned, the latter in case of three or more. In the case of the DL methods (RNN, LSTM, GRU and CNN1D), a randomized hyperparameter tuning was performed to search for the most optimal set of parameters: number of neurons, batch size, sequence length and number of training epochs without improving the validation metric (referenced as patience in the table).

4.5. Results

Three experiments have been carried out for each of the datasets. First, the 12 ML/DL models were trained with all the available predictor variables, the results of which can be found in Tables 8, 10 and 12. Next, a sensitivity analysis was performed in order to examine the influence of each group of variables (Fig. 10). For this, each of the 12 methods was rerun four times, with one group of variables removed in each run. The influence of different sets of variables upon the performance of the regression methods can be seen in Figs. 15, 16, 18, 19, 21 and 22. Finally, the FS algorithm was executed with each of the regression techniques. The results are shown in Tables 9, 11 and 13. Also, a time-series figure with the comparison of the actual number of users and the prediction of the best model for each dataset is attached in Figs. 17, 20 and 23.

4.5.1. BiciMad

Results for the BiciMad dataset when using all the explanatory variables are presented in Table 8. It can be observed that best goodness-of-fit (R^2) was performed by both SVR and ELM, indicating that these two methods explained the most of the dependent variable's variance. The lowest RMSE was scored by the ELM, although no significant difference was observed with the SVR and CNN. Generally, RMSE is an indicator similar to variance, quantifying the deviations in predicted and actual values squared. Distinction of the RMSE towards e.g. MAPE or MAE is that the former penalizes outliers to a much higher degree (by a square function). All regressors (LR, PR and Lasso) on one hand score exceptionally well MAPEs and MAEs, but perform poorly on RMSEs, indicating that severe outliers are supposed to be present in linear predictions. However, in linear error terms these outliers do not gain such importance and consequently result is among the best.

Subsequently, sensitivity analyses on a BiciMad dataset for all ML/DL methods were run (Figs. 15 and 16). In average, it may be observed that all groups of variables are relevant to solve the regression problem, since the removal of any group of features worsens performance metrics (R^2 decreases and the rest of indexes increases). In addition, it is worth mentioning that the variables related to users of previous days (persistence-related variables) play an essential role in the regression models, being particularly critical in ML models, since removing these variables dramatically worsens the results, making the R^2 below 0.5 in most cases and almost doubling the values of the rest of the metrics.

Next, the FS method was executed, obtaining the optimal feature set that minimizes the training MAE error for each model (Table 9), where 1 indicates that the corresponding variable is being used and 0 that remains unused. Employed regression models find a very similar optimal subset after the FS, where only a single feature, i.e., Month label, is removed in 4 of the 12 cases. Also, the last two digits of the feature combination vector denotes the representation way of Day-of-Week and Month label (0 = numeric, 1 = categorical dummy variables), where disparity between methods is observed.

It may be appreciated that statistical indicators covered do not change much with respect to the values showed previously without the FS process, with some methods with a significant improvement (MLP, LSTM or GRU) and some others which perform worse than previously (PR, RF, SVR and CNN). This can be explained by considering that, in first place, some complex models do not need careful feature engineering and can automatically discover useful features from high-dimensional raw data, and second, the best set of features is selected

Table 6
Correlation coefficients between the target and predictor variables for the Bicing Barcelona dataset, sorted in descending correlation order.

Predictor variables	R	Predictor variables	R
AverageAnalogHours	0.8724	Previous12HourAverage	0.2706
LastAnalogHour	0.8464	SameDayAverageTemperature	0.2624
LastAnalogDay	0.7920	SameDayPreviousWeek	0.2485
PreviousMonthSameHourAverage	0.7789	SameDay4WeeksAverage	0.2370
PreviousWeekSameHourAverage	0.7494	SameDay4WeeksAverage	0.1842
Previous2HourAverage	0.7174	PreviousMonthAverage	0.1765
PreviousHourAverage	0.6982	PreviousDaySameHourTemperature	0.0945
Previous4HourAverage	0.6132	Month	0.0934
PreviousDaySameHourAverage	0.6106	NextDayAverageTemperature	0.0923
SameDaySameHourPreviousWeek	0.5992	PreviousDayAverageTemperature	0.0475
Hour	0.4780	NextDay_DayofWeek	0.0435
AverageAnalogDays	0.4063	NextDayHoliday	−0.0386
Temperature	0.4002	Holiday	−0.1092
PreviousDayAverage	0.2906	DayofWeek	−0.2119
SameDaySameHour4WeeksAverage	0.2830		

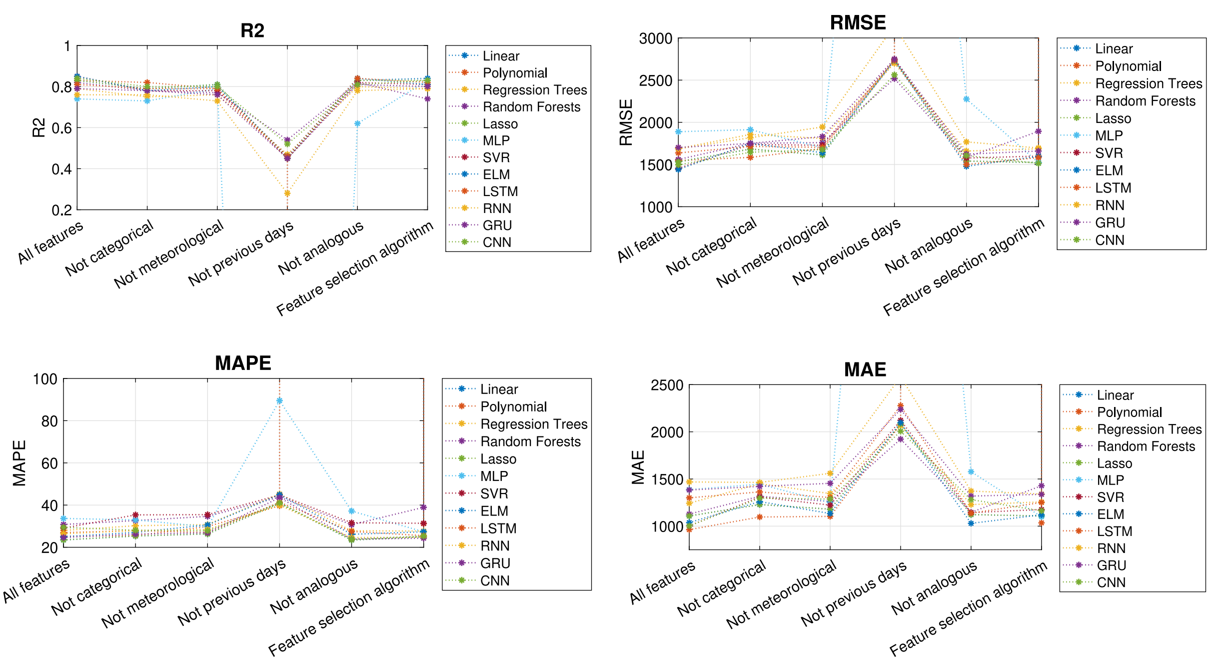


Fig. 15. Test error metrics for each model tested with BiciMad dataset and different features.

based on the train error, which does not necessarily correlate with the best test error, causing some models to become too specific. Overall, the best assessments change for the R^2 and RMSE indicators, the LR works the best regarding the R^2 indicator and the CNN works the best at RMSE. Fig. 16 shows the comparison for the error indexes when applying the FS algorithm and without it for the ML and DL methods, exhibiting a slight worsening in the test metrics for the case of ML models, mainly due to poor RF performance after FS: R^2 decays by a 1.22%, RMSE by a 2.35%, MAPE by a 6.93% and MAE by a 3.94%. For the DL methods, the four metrics improve when applying the FS algorithm: R^2 improves by a 1.25%, RMSE by a 1.09%, MAPE by a 7.00% and MAE by a 1.21%.

Finally, an out-of-sample time series prediction for the RMSE-best ML/DL method is visualized graphically in Fig. 17, consisting of both the actual and predicted values. No large deviations or outliers are detected in general.

4.5.2. Madrid Cable Car

Next, results of the ML methods in the Madrid Cable Car dataset are reported in Table 10 when using all the predictor variables (15) as inputs for the models. The PR model performance is not shown due to extremely poor results. Among the rest of the ML techniques applied, consistency is observed in the four error metrics, with low dispersion for all the models, excluding RT, ELM and CNN, which performs much worse than the others. Both the best coefficient of determination R^2 and RMSE were scored by the SVR method and the best MAPE and MAE by LSTM/RNN. It should be recalled that this was the most complex database, as it had a high percentage of days without data collection, also, it is the problem with the lowest number of training instances, 342, so it is reasonable to expect that the results are worse than in the other two datasets, with lower correlation coefficients. In fact, as the continuity of the time series is lost due to the presence of incomplete instances, time-series specialized DL models do not perform as optimally as in the other cases, with the CNN model remaining lost.

Sensitivity analyses for the Madrid cable car may be observed in Figs. 18 and 19, and they are consistent for all included statistical indicators. Two groups of variables critically affect overall statistical performance, i.e., the categorical and analogous predictor variables, meaning that excluding either of the two from dataset results in a significant drop of R^2 and a simultaneous significant increase of RMSE, MAPE and MAE. On the other hand, meteorological variables (contrary to expected) do not significantly affect the overall statistical performance, instead an improvement of R^2 is observed by excluding this group of variable. Similarly, exclusion of the predictor variables

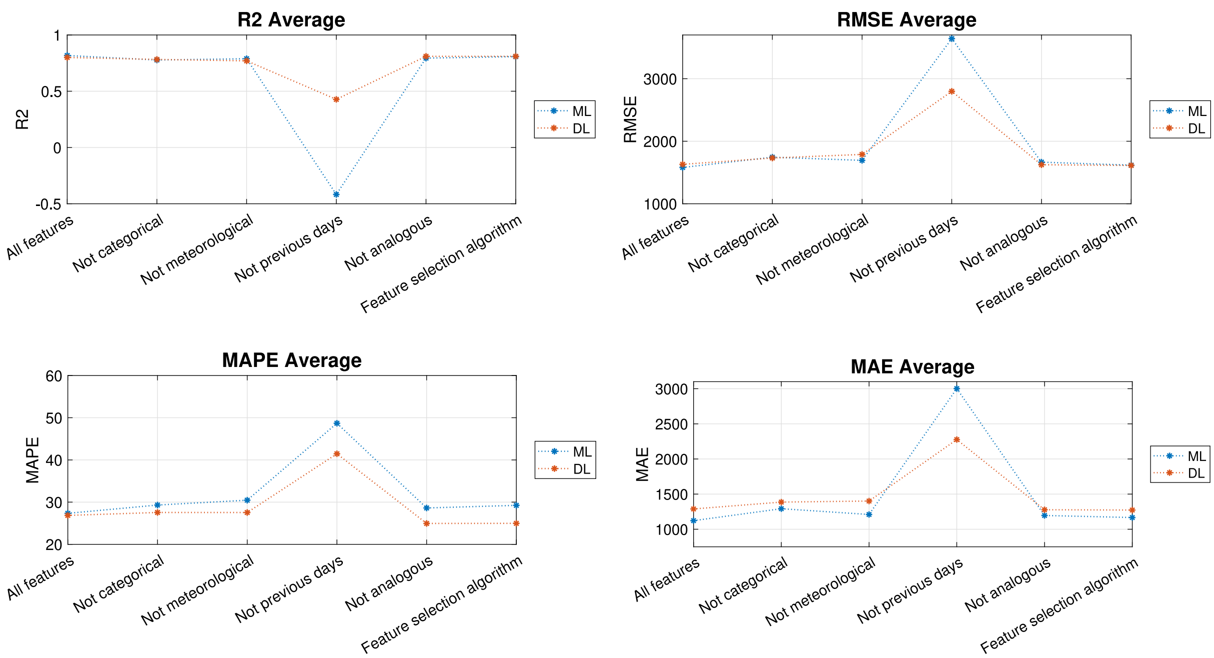


Fig. 16. Test error metrics average for BiciMad dataset and different features.

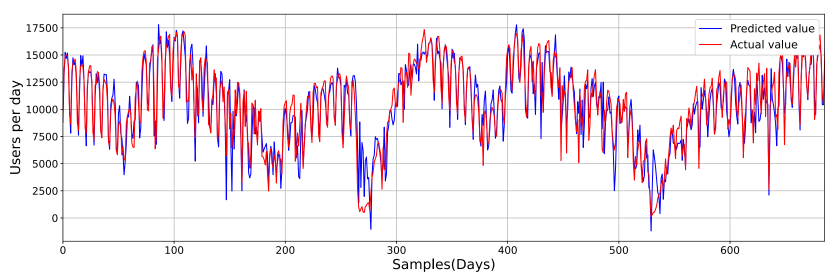


Fig. 17. BiciMad time series prediction for the ELM model.

on previous days does not significantly harm performance, instead improved MAPE and MAE indicators are observed, this might have been expected since, having so many days without data, only the previous week's and month's averages are taken into account.

Finally, results after FS are reported in Table 11 along with the optimal feature combination selected for each model. The number of fixed predictor variables has been set to 9, that corresponds to the 60% of the most correlated variables towards the target. For the rest of the predictors, it is observed that each model prefers a different parameter combination, which relates to the one providing the lowest training error. The last two digits of the best feature combination vector refer to the representation form of the variables DayofWeek and Month (0 = numeric, 1 = categorical dummy variables), it may be observed how for the day-of-week label all methods except RT prefer the option of using dummy variables, while for the month label MLP, ELM and RNN are the exception.

Regarding the performance of ML regression methods after executing the FS algorithm, there are overall improvements of all methods in MAPE and MAE metrics, with the exception of RF, which performs slightly worse. For the R^2 and RMSE, LR, RT and RF performs worse than previously, but a significant improvement is observed in the rest of methods reaching a R^2 of 0.75 for SVR. It must be reminded that best feature combination is selected for minimizing the training MAE error with the partial exhaustive search, therefore it may occur that a better training error does not relate to a better performance on the test set. Even so, a general improvement of the regression error is exhibited after applying the FS algorithm, as it is reported in Fig. 19, where an enhancement of the averages of the 7 ML methods (PR has been exclude to avoid tampering) for the 4 error metrics is appreciated: R^2 improves by a 25.39%, RMSE by a 3.88%, MAPE by a 9.61% and MAE by a 8.23%.

In the case of the DL models, it is observed that for the RNN and CNN methods, the use of FS leads to a significant improvement of the 4 error metrics, while for LSTM it results in a deterioration of the performance. GRU achieves similar results in both cases. Average improving rates are: −1.88% for R^2, 3.52% for RMSE, 4.63% for MAPE and 2.78% for MAE.

RNN model exhibits the best performance with the best results in all error indexes. Time prediction for test set inferred with this model is depicted in Fig. 20.

4.5.3. Bicing Barcelona

Results for the third regression problem, Bicing Barcelona, when inputting all the features available are reported in Table 12. Here, large differences in results between ML/DL methods are found. Again, PR results are omitted due to disastrous performance. First, three negative R^2 assessments are observed, i.e., for ELM, LR and SVR, which performs very poorly, even though all of them were among the better ML/DL methods in Bicimad predictions and LR and SVR in cable car dataset. These methods are very poor according to the rest of the indicators as well. For the given dataset, CNN method is an obvious choice, outperforming all the other methods significantly in the four metrics. Also, CART methods performs exceptionally well too. The advantage of the RF is that it is an ensemble-based, which means that not a

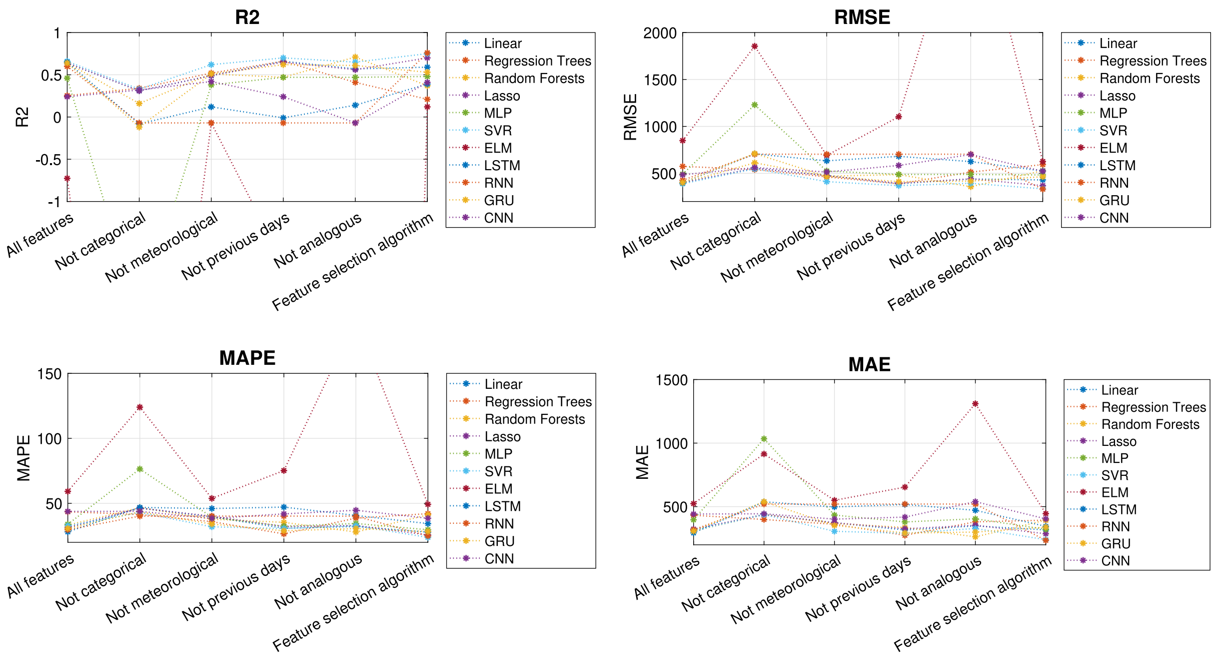


Fig. 18. Test error metrics for each model tested with Madrid Cable Car dataset and different features.

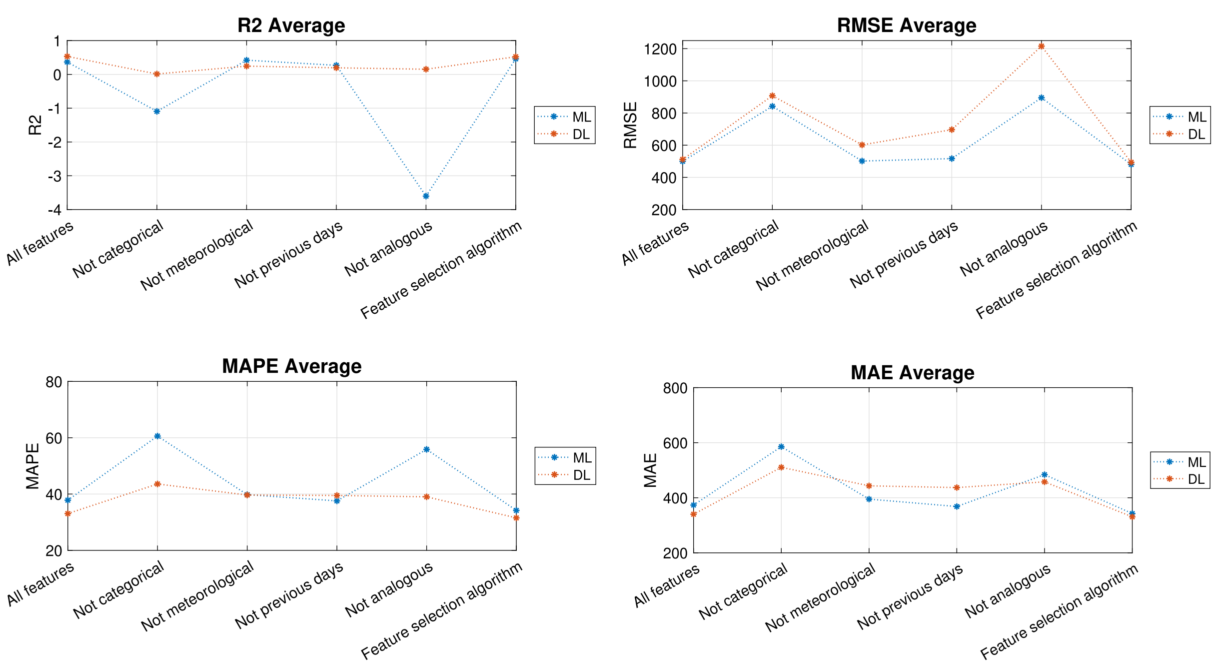


Fig. 19. Test error metrics average for Madrid Cable Car dataset and different features.

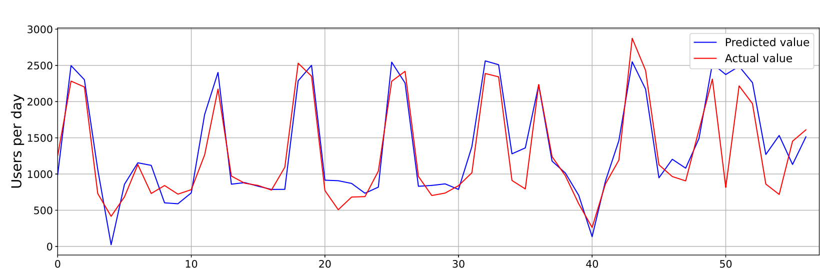


Fig. 20. Madrid's cable car time series prediction for the RNN model.

Table 7
Experimental setup.

Lasso		SVR	
Alpha	0.1,0.2,…,4.9,5	C	10,110,210, …810,910
Search	Grid	Epsilon	1e−5,.0112,.0222,…, .0889,.1
		Search	Grid
RF		**RT**	
No. of estimators	1–11	Max. depth	1,2,…,19,20
Max. Features	2–21	Min. leaf	1,2,…,19,20
Bootstrap	True, False	Search	Grid
Search	Randomized		
MLP		**LSTM**	
N_{inp}	No. of features	N_{LSTM}	16,32,64,128,256
N_{hid1}	40–60	N_{hid}	16,32,64,128
N_{hid2}	20–40	N_{out}	1
N_{out}	1	Patience	50,75,100,125
Activation	relu, sigmoid	Batch size	16,32,64
Epochs	500	Sequence	2,3,4,5,6
Batch size	20	Search	Randomized
Search	Randomized		
RNN		**GRU**	
N_{RNN}	16,32,64,128,256	N_{GRU}	16,32,64,128,256
N_{hid}	16,32,64,128	N_{hid}	16,32,64,128
N_{out}	1	N_{out}	1
Patience	50,75,100,125	Patience	50,75,100,125
Batch size	16,32,64	Batch size	16,32,64
Sequence	2,3,4,5,6	Sequence	2,3,4,5,6
Search	Randomized	Search	Randomized
CNN1D			
N_{CNN1D}	16,32,64,128,256		
N_{hid}	16,32,64,128		
N_{out}	1		
Patience	50,75,100,125		
Kernel	2,3,4,5,6		
Sequence	2,3,4,5,6		
Batch size	16,32,64		
Search	Randomized		

Table 8
Performance of the evaluated ML regression methods for the BiciMad dataset when including all available predictor variables. Best assessments in bold.

Method	R^2	RMSE	MAPE (%)	MAE
LR	0.83	1531.90	**23.48**	1107.60
PR	0.83	1547.48	25.19	**963.09**
RT	0.79	1688.84	27.99	1243.04
RF	0.82	1559.27	30.80	1128.09
Lasso	0.83	1531.89	**23.48**	1107.89
MLP	0.74	1888.60	33.66	1388.06
SVR	**0.85**	1454.20	29.09	1002.48
ELM	**0.85**	**1441.00**	25.04	1037.89
LSTM	0.81	1637.39	26.93	1300.33
RNN	0.76	1686.16	26.40	1468.65
GRU	0.79	1702.02	24.63	1383.02
CNN	0.84	1496.25	29.45	1002.79

single model is built, but rather a whole spectre of them, compared to the rest of the ML/DL methods that only have a single chance. Still, one must note that a single element of RF, the RT performs exceptionally well too. RTs have a unique property compared to the rest of the ML/DL methods, i.e., they make predictions based on decision boundary placements, meaning that a different (but similar) sample of input datapoints may have the same (averaged) predicted value. In this way, predicting outliers is much suppressed, indicating better statistical assessments. The rest of the methods behave in a similar way without notable differences.

Regarding the sensitivity analysis for this dataset, which is depicted in Figs. 21 and 22, it is worthwhile to mention the negative influence of the presence of categorical features in the performance of the regression methods, making the four error metrics to be significantly better when removing them. Concerning the other three groups of variables, it seems that they are meaningful for the implemented ML/DL methods, since removing them deteriorates the error metrics, with a special relevance of the meteorological variables, and analogous for the case of ML models.

Table 9
Performance of the evaluated ML regression methods for the BiciMad dataset after running the feature selection algorithm. Best assessments in bold.

Method	R^2	RMSE	MAPE (%)	MAE	Best feature combination
LR	**0.84**	1522.85	25.01	1107.44	1111111111111111011111
PR	0.81	1596.68	31.19	**1034.83**	1111111111111111011111
RT	0.79	1686.01	27.68	1256.53	1111111111111001010000
RF	0.74	1893.59	38.93	1428.54	1111111111110000000011
Lasso	0.83	1522.89	25.02	1107.71	1111111111111111011111
MLP	0.83	1522.14	27.33	1099.77	1111111111111101111010
SVR	0.81	1599.43	31.38	1176.43	1111111111111111011111
ELM	0.81	1597.04	27.35	1120.63	1111111111111111111100
LSTM	0.82	1581.97	25.52	1251.14	1111111111111111110000
RNN	0.79	1695.46	24.75	1340.74	1111111111111111011010
GRU	0.80	1659.60	**24.31**	1337.07	1111111111111100111001
CNN	0.83	**1513.58**	25.31	1163.39	1111111111111010110000

Table 10
Performance of the evaluated ML regression methods for the Madrid Cable Car dataset when including all available predictor variables. Best assessments in bold.

Method	R^2	RMSE	MAPE (%)	MAE
LR	0.64	400.59	30.71	318.12
RT	0.25	573.96	43.50	432.43
RF	0.64	396.32	33.22	308.70
Lasso	0.64	398.72	31.49	323.72
MLP	0.46	490.70	33.86	396.46
SVR	**0.66**	**389.07**	32.80	310.45
ELM	−0.73	848.78	59.19	523.83
LSTM	0.65	397.27	**28.08**	**293.30**
RNN	0.60	426.29	29.67	318.37
GRU	0.64	401.24	30.73	307.11
CNN	0.24	484.43	43.78	441.81

Table 11
Performance of the evaluated ML regression methods for the Madrid Cable Car dataset after running the feature selection algorithm. Best assessments in bold.

Method	R^2	RMSE	MAPE (%)	MAE	Best feature combination
LR	0.59	429.51	27.82	317.05	11111111111101011
RT	0.21	592.94	42.04	393.52	11111111111110101
RF	0.37	528.76	41.11	394.10	11111111100001011
Lasso	0.70	368.47	25.66	286.53	11111111111101111
MLP	0.48	482.79	29.67	321.90	11111111100100110
SVR	0.75	334.43	23.80	240.47	11111111111011011
ELM	0.12	625.57	49.23	445.12	11111111110111010
LSTM	0.39	524.99	34.39	345.14	11111111100010011
RNN	**0.76**	**331.21**	**24.95**	**234.67**	11111111110011110
GRU	0.53	463.46	28.30	339.14	11111111110111001
CNN	0.41	522.78	38.49	403.79	11111111111110011

Table 12
Performance of the evaluated ML regression methods for the Bicing Barcelona dataset when including all available predictor variables. Best assessments in bold.

Method	R^2	RMSE	MAPE (%)	MAE
LR	−0.08	182.63	73.87	149.06
RT	0.76	85.58	37.56	59.03
RF	0.86	66.58	26.75	43.49
Lasso	0.35	141.03	63.25	114.98
MLP	0.32	145.06	49.93	110.79
SVR	−0.27	197.75	280.22	161.94
ELM	−12.75	619.72	373.83	361.51
LSTM	0.35	140.69	53.33	100.45
RNN	0.57	114.72	68.44	79.18
GRU	0.61	107.01	38.10	67.94
CNN	**0.89**	**58.15**	**26.23**	**40.08**

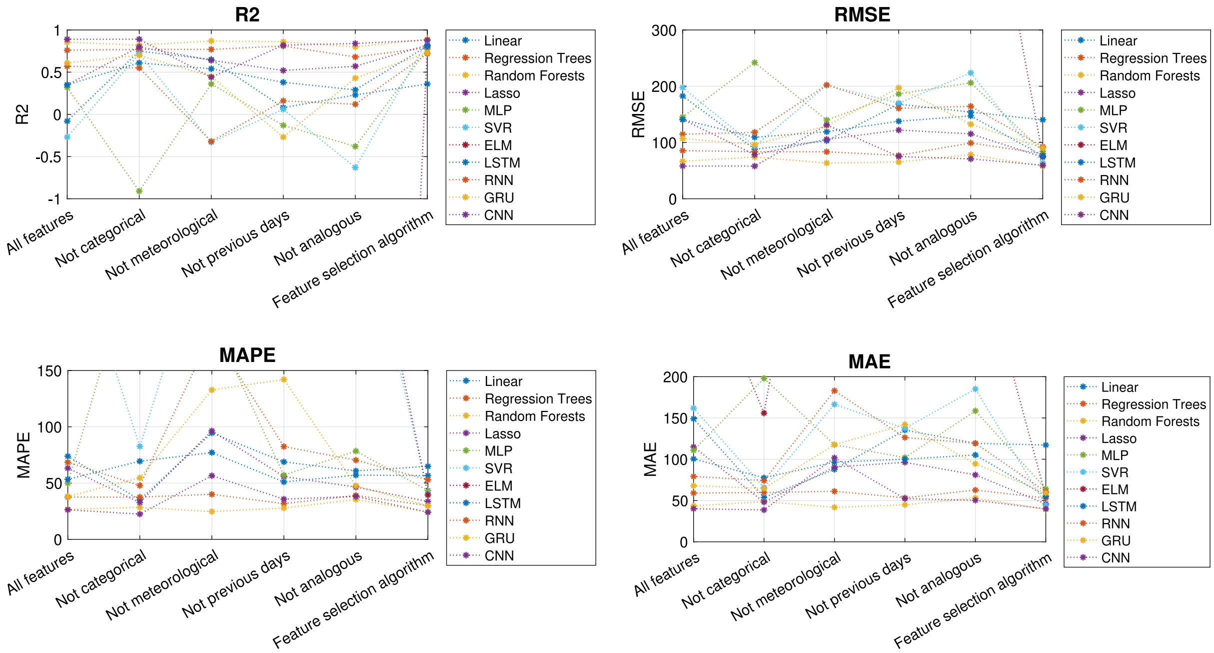


Fig. 21. Test error metrics for each model tested with Bicing Barcelona dataset and different features.

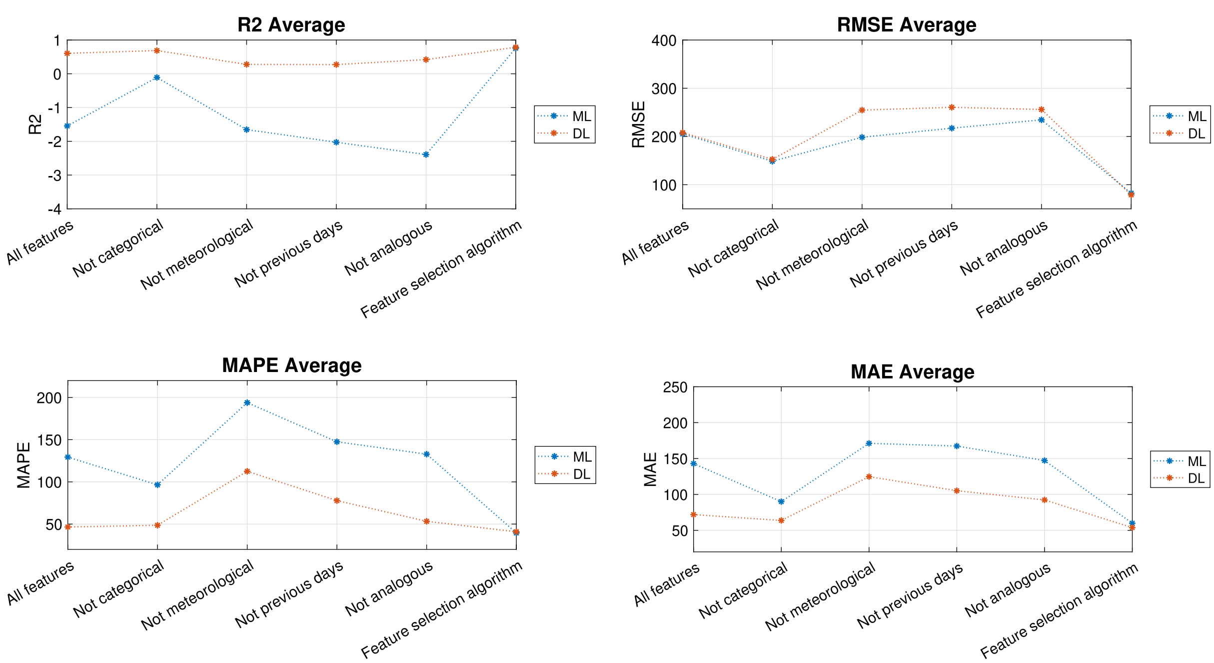


Fig. 22. Test error metrics average for Bicing Barcelona dataset and different features.

Excellent results were achieved after applying the FS algorithm in this dataset. Results are reported in Table 13, excluding the PR which achieved disastrous results again. It may be appreciated how all the methods improved for every metric, resulting in all of them having a very similar performance with the exception of LR which remains a bit lost. In this case, three variables were tested both as numerical and categorical dummy variables: hour, day-of-week and month label. All the methods but the RT, LSTM and RNN preferred the hour variable as 24 categorical dummy variables, while some disparity appears with the other two features tested.

Fig. 22 plots the average error metrics with and without FS for the ML and DL models separately, exhibiting a huge improvement in all of them: in the case of ML models, R^2 improves by an impressive 149.09%, RMSE by a 59.92%, MAPE by a 69.48% and MAE by a 58.12%; and in the case of DL models, R^2 improves by a 29.75%, RMSE by a 62.06%, MAPE by a 12.11% and MAE by a 25.22%. This radical improvement can be attributed to the fact that this was the database with the largest number of predictor variables (29), so prior to applying FS, problems of redundancy or multi-collinearity could be occurring.

Time series prediction inferred with the RF model after applying the FS algorithm is depicted in Fig. 23.

5. Discussions and conclusions

The ability to anticipate the ups and downs in the demand for public services is essential for the proper management of mobility issues in

Table 13
Performance of the evaluated ML regression methods for the Bicing Barcelona dataset after running the feature selection algorithm. Best assessments in bold.

Method	R^2	RMSE	MAPE (%)	MAE	Best feature combination
LR	0.36	140.41	64.90	117.20	11111111111111111111111111101111110
RT	0.79	80.16	29.97	52.96	1111111111111111110011100010000000
RF	**0.89**	**57.91**	24.16	**39.28**	1111111111111111110000000000010100
Lasso	0.82	74.01	33.61	45.95	1111111111111111111111100111101111
MLP	0.77	84.31	43.31	64.12	1111111111111111111111111000111110
SVR	0.87	63.58	40.94	44.54	1111111111111111111111110110001111
ELM	0.81	76.09	39.40	55.07	1111111111111111110000001000000110
LSTM	0.81	76.28	56.63	55.47	1111111111111111111111110110100001
RNN	0.72	92.45	52.97	60.61	1111111111111111111111111010010000
GRU	0.73	90.24	29.85	59.12	1111111111111111111111110010001100
CNN	0.88	59.66	**24.14**	39.90	1111111111111111111111111010001100

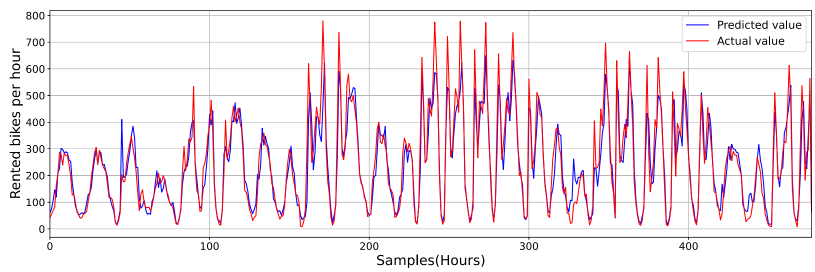


Fig. 23. Bicing Barcelona time series prediction for the random forest model.

cities. For example, it would be possible to allow the reinforcement of certain forms of public transport on peak demand dates or providing different alternatives to the population, when necessary. All this contributes to the development and sustainability of green mobility systems, making these systems competitive and capable of responding to the needs of the population. In addition, being able to adapt the offer of these transport systems to the respective demand, enables savings in both economic costs and environmental pollution.

Three regression problems related to Green Mobility in big cities of Spain have been tackled in this paper, with ML and DL regression techniques. Specifically, the bike sharing demand in Madrid and Barcelona, and number of users of Madrid's cable car were predicted. The time-sequence datasets analyzed in this study were challenging. The time series of bike sharing demand are not likely stationary. Instead, a gradual increase (BiciMad) or decrease (Bicing Barcelona) are observed. It is then of crucial importance that the model receives the information of past demand, otherwise significant deviations and outliers may be expected. Also, working with human activity-related datasets implies special characteristic of the problems, involving different kind of properties and peculiarities for every case. An essential issue to take into account when solving this kind of timed-series predictions is the database construction with the addition of exogenous variables. Four groups of predictors have been categorized according to the information provided by each one of them: categorical, meteorological, previous days information and analogous instances (a novelty of this work). A sensitivity analysis was performed in each problem to examine the importance of every group. It was concluded that this is a specific characteristic of each problem, since, depending on each dataset, the results improve or got worse when removing a particular group of variables. For the case of the BiciMad database, it was found that the persistence-related variables remain as the most relevant ones. In the case of Madrid Cable Car, the calendrical/categorical variables, together with the analogous instances predictors, were the most important ones. Finally, in the case of Bicing Barcelona, the categorical variables proved to have a negative impact on most of the models, while the meteorological features were the most relevant ones.

The main novelties and contributions presented in this work, with respect to the current state of the art, can be outlined as follows: first, this is the first time that the problem of passenger flow forecasting on a cable car have been addressed using ML and DL multivariate techniques with time-series exogenous variables, achieving excellent results despite the difficulties associated with the discontinued database. Second, the inclusion of an additional group of predictive variables related to the information regarding the demand data of analogous instances to the current prediction, allowing to take into account similarities between the current and past events. This set of features has been found to be significantly associated with the target variable for the three databases considered, both in terms of the correlation coefficients as well as the significant performance degradation suffered by the prediction models when removing this set of predictor variables, especially in the case of the prediction of cable car passengers. Third, an exhaustive analysis of which sets of features are most relevant for each prediction method is presented. We conclude that, on the one hand, relations among predictors corresponding to each group show similar patterns in all the databases, as well as regarding the target variable, where analogous and persistence related variables are the ones showing higher correlation coefficients; on the other hand, results indicate that the specific problems need to address the selection of features that best fits to each regression model individually.

Upon a closer examination of the results, it can be concluded that predictions made for the three problems studied are reasonably reliable, with correlation coefficients (between the actual and predicted values) in the range of 0.76 to 0.89, and relative errors below 25%. The obtained findings demonstrate that bicycle usage databases deliver the most favorable outcomes in terms of prediction performance, while in the case of the Madrid cable car, the results are the poorest among all the databases considered, yet showing satisfactory error metrics. It is important to highlight that this specific database contained numerous incomplete samples and incorporated the fewest exogenous variables among the datasets considered, so these outcomes were somehow expected. In the time series prediction graphs (Figs. 17, 20 and 23) it may be appreciated how the models correctly anticipate the trend of the data for all the studied cases, predicting successfully the days in which the demand increases or decreases. Although some outliers or unexpected values appear that the system is not able to correctly forecast, it must be kept in mind that these are human activity related problems, and therefore an stochastic component is always present. In addition, the number of predictors variables taken into account is not too large, and there may be some important data not contemplated, e.g., pedestrian mobility in big cities can be affected by major events taking place in the city, such as sports matches, music concerts or political elections, which have not been considered in this paper.

Also, the results shown in this paper indicate the importance of selecting the most appropriate features for each problem, since these are very specific problems that need to be addressed individually. For this reason, it was decided to apply a FS algorithm that enables to automatically select the optimal predictor variables for each problem and for each ML/DP model used. Great improvements in prediction performance were observed after the application of the FS algorithm for two out of the three datasets, reducing the average test MAE error for the ML by 8.23% in the Madrid Cable Car dataset and by 58.12% in the Bicing Barcelona dataset. When the initial predictor variables have similar levels of significance, the results remain practically constant slightly worsening for the BiciMad dataset (test MAE decays by a 3.94%). This is due to the fact that minimizing training errors in some cases may lead to highly specific models, thus worsening test errors. Regarding the effect of the application of FS in the DL models, it is observed that the improvement rates obtained are lower than in the ML cases, due to the fact that these complex models already perform an extraction of the most important parameters, and thus they are less sensitive to the previous data treatment or to the overdimensionality of the data. Average test MAE errors improve by a 2.78% in the Madrid Cable Car dataset and by 25.22% in the Bicing Barcelona dataset, in the case of BiciMad dataset, MAE on test data improves by a 11.21% when applying the FS algorithm. In terms of the specific performance of the multivariate regression techniques employed, it has been observed that the specialized DL models for time series treatment perform satisfactorily in all cases, with similar performance for the four methodologies tested. In addition, among the ML methodologies, RF and SVR perform very robustly, achieving very competitive results in all cases. The use of 12 different prediction methods (8 from ML and 4 from DL) has served two purposes: (1) allow the comprehensive extraction of global conclusions about the importance of predictor variables, observing common trends for all the studied cases (Figs. 15, 18 and 21); and (2) being the performance of each method case specific, the comparison among a large number of model enables to select the best performing methods for each database, so when implementing these models in operation, the idea would be to select the 2–3 best models that work best in each case and use them as demand prediction methods, so that they might be useful to support logistic decisions related to green mobility in large cities.

Further lines of work will be focused on three main aspects. First, the search for new predictor variables that, a-priori, may be unrelated to the problem, but can have a major effect on the results, e.g., major events taking place in the city, or additional meteorological variables such as wind or humidity data. Second, extend the predictions for more than one day ahead, thus city logistics can be scheduled and managed several days or weeks in advance. This extension of the prediction time-horizon may be problematic, in the sense that the accuracy of the models will reduce. Thus, the analysis of this performance reduction is basic in order to extend the proposed ML models to longer prediction time-horizons. Finally, the application of the described and implemented ML methods to new databases related to the use of green mobility services in big cities, e.g., subway, buses, scooter rentals, urban car rentals, etc. is an exciting possibility. This framework could be of help to city councils' traffic management sections, allowing them to anticipate the mobility of the population, and to adapt public services to their needs. The implemented ML algorithms are designed to be flexible, requiring only the construction of the database with significant exogenous variables when tackling a new prediction problem.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This research has been partially supported by the project PID2020-115454GB-C21 of the Spanish Ministry of Science and Innovation (MICINN). This research has been partially funded by the Autonomous Community of Madrid through ELLIS consortium.

References

Ai, Y., Li, Z., Gan, M., Zhang, Y., Yu, D., Chen, W., et al. (2019). A deep learning approach on short-term spatiotemporal distribution forecasting of dockless bike-sharing system. *Neural Computing and Applications*, *31*(5), 1665–1677.

Alessandrini, S., Delle Monache, L., Sperati, S., & Cervone, G. (2015). An analog ensemble for short-term probabilistic solar power forecast. *Applied Energy*, *157*, 95–110.

Alessandrini, S., Delle Monache, L., Sperati, S., & Nissen, J. (2015). A novel application of an analog ensemble for short-term wind power forecasting. *Renewable Energy*, *76*, 768–781.

Alzaman, C., Aljuneidi, T., & Li, Z. (2023). Predicting bike usage and optimizing operations at repair shops in bike sharing systems. *IEEE Access*.

Asencio-Cortés, G., Florido, E., Troncoso, A., & Martínez-Álvarez, F. (2016). A novel methodology to predict urban traffic congestion with ensemble learning. *Soft Computing*, *20*(11), 4205–4216.

Ashqar, H. I., Elhenawy, M., Almannaa, M. H., Ghanem, A., Rakha, H. A., & House, L. (2017). Modeling bike availability in a bike-sharing system using machine learning. In *2017 5th IEEE international conference on models and technologies for intelligent transportation systems (MT-ITS)* (pp. 374–378). IEEE.

Ayuntamiento de Barcelona (2022). Uso del servicio Bicing de la ciudad de Barcelona. https://datos.gob.es/es/catalogo/l01080193-uso-del-servicio-bicing-de-la-ciudad-de-barcelona, Online; accessed 09 March 2022.

Ayuntamiento de Madrid (2022a). Bicimad. Alta de usuarios y usos por día del servicio público de bicicleta eléctrica. https://datos.gob.es/es/catalogo/l01280796-bicimad-alta-de-usuarios-y-usos-por-dia-del-servicio-publico-de-bicicleta-electrica1, Online; accessed 09 March 2022.

Ayuntamiento de Madrid (2022b). Teleférico de madrid. número de usuarios. https://datos.gob.es/es/catalogo/l01280796-teleferico-de-madrid-numero-de-usuarios1, Online; accessed 09 March 2022.

Bassolas, A., Barbosa-Filho, H., Brian, D., Dotiwalla, X., Eastham, P., Gallotti, R., et al. (2019). Hierarchical organization of urban mobility and its connection with city livability. *Nature Communications*, *10*(4817), 41–51. http://dx.doi.org/10.1038/s41467-019-12809-y.

Benchimol, M., Benchimol, P., Chappert, B., De La Taille, A., Laroche, F., Meunier, F., et al. (2011). Balancing the stations of a self service "bike hire" system. *RAIRO-Operations Research*, *45*(1), 37–61.

Bi, F., He, T., Xie, Y., & Luo, X. (2023). Two-stream graph convolutional network-incorporated latent feature analysis. *IEEE Transactions on Services Computing*.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.

Çolak, S., Lima, A., & González, M. C. (2016). Understanding congested travel in urban areas. *Nature Communications*, *7*.

Charandabi, N. K., Gholami, A., & Bina, A. A. (2022). Road accident risk prediction using generalized regression neural network optimized with self-organizing map. *Neural Computing and Applications*, *2022*, 1–14. http://dx.doi.org/10.1007/s00521-021-06549-8.

Chen, P.-C., Hsieh, H.-Y., Su, K.-W., Sigalingging, X. K., Chen, Y.-R., & Leu, J.-S. (2020). Predicting station level demand in a bike-sharing system using recurrent neural networks. *IET Intelligent Transport Systems*, *14*(6), 554–561.

Cheng, J., Dong, L., & Lapata, M. (2016). Long short-term memory-networks for machine reading. arXiv preprint arXiv:1601.06733.

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555.

Collini, E., Nesi, P., & Pantaleo, G. (2021). Deep learning for short-term prediction of available bikes on bike-sharing stations. *IEEE Access*, *9*, 124337–124347.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297.

Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F. R., Gaughan, A. E., et al. (2014). Dynamic population mapping using mo-bile phone data. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, 15888–15893. http://dx.doi.org/10.1073/pnas.1408439111.

Ding, S., Xu, X., & Nie, R. (2014). Extreme learning machine and its applications. *Neural Computing and Applications*, *25*(3), 549–556.

Dokuz, A. S. (2021). Fast and efficient discovery of key bike stations in bike sharing systems big datasets. *Expert Systems with Applications*, *172*, Article 114659.

Eubank, S., Guclu, H., Kumar, V. S. A., Marathe, M. V., Srinivasan, A., Toroczkai, Z., et al. (2004). Modelling disease outbreaks in realistic urban social networks. *Nature*, *429*, 180–184.

Fister, D., Perc, M., & Jagrič, T. (2021). Two robust long short-term memory frameworks for trading stocks. *Applied Intelligence*, *51*(10), 7177–7195.

Fukushima, K., & Miyake, S. (1982). Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In *Competition and cooperation in neural nets* (pp. 267–285). Springer.

Gardner, M. W., & Dorling, S. (1998). Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric Enviroment*, *32*(14–15), 2627–2636.

Garsous, G., Suárez-Alemán, A., & Serebrisky, T. (2019). Cable cars in urban transport: travel time savings from La Paz-El Alto (Bolivia). *Transport Policy*, *75*, 171–182.

Gauss, C. F. (1823). *Theoria combinationis observationum erroribus minimis obnoxiae, vol. 2*. H. Dieterich.

Graves, A. (2013). Generating sequences with recurrent neural networks. arXiv preprint arXiv:1308.0850.

Graves, A., Mohamed, A.-r., & Hinton, G. (2013). Speech recognition with deep recurrent neural networks. In *2013 IEEE international conference on acoustics, speech and signal processing* (pp. 6645–6649). Ieee.

Guizzardi, A., Pons, F. M. E., Angelini, G., & Ranieri, E. (2021). Big data from dynamic pricing: A smart approach to tourism demand forecasting. *International Journal of Forecasting*, *37*(3), 1049–1060.

Guzman, L. A., Cantillo-Garcia, V. A., Arellana, J., & Sarmiento, O. L. (2022). User expectations and perceptions towards new public transport infrastructure: evaluating a cable car in Bogotá. *Transportation*, 1–21.

Harikrishnakumar, R., & Nannapaneni, S. (2023). Forecasting bike sharing demand using Quantum Bayesian Network. *Expert Systems with Applications*, Article 119749.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049.

Ho, T. K. (1995). Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition, vol. 1* (pp. 278–282). IEEE.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. http://dx.doi.org/10.1162/neco.1997.9.8.1735.

Hofer, K., Haberl, M., & Fellendorf, M. (2017). Travel demand estimation of cable cars supplementing public transport. In *HEART 2017: 6th Symposium of the european association for research in transportation* (pp. 1–3).

Hofer, K., Haberl, M., Fellendorf, M., Huber, G., & Fallast, K. (2018). Travel demand estimation for cable car transport in the urban areas shown for the moderate-sized city of graz, Austria. In *Presented at transport research arena 2018: A digital era for transport* (pp. 1–10).

Hoffmann, K., & Liehl, R. (2005). *Cable-drawn urban transport systems, vol. 77*. WIT Press.

Hua, M., Chen, J., Chen, X., Gan, Z., Wang, P., & Zhao, D. (2020). Forecasting usage and bike distribution of dockless bike-sharing using journey data. *IET Intelligent Transport Systems*, *14*(12), 1647–1656.

Huang, D., Ahmed, A., Arafat, S. Y., Rashid, K. I., Abbas, Q., & Ren, F. (2020). Sentence-embedding and similarity via hybrid bidirectional-lstm and cnn utilizing weighted-pooling attention. *IEICE TRANSACTIONS on Information and Systems*, *103*(10), 2216–2227.

Huang, S., Cai, N., Pacheco, P. P., Narrandes, S., Wang, Y., & Xu, W. (2018). Applications of support vector machine (SVM) learning in cancer genomics. *Cancer Genomics & Proteomics*, *15*(1), 41–51.

Huang, G.-B., Zhu, Q.-Y., & Siew, C.-K. (2006). Extreme learning machine: theory and applications. *Neurocomputing*, *70*(1–3), 489–501.

Islam, M. J., Datta, R., & Iqbal, A. (2023). Actual rating calculation of the zoom cloud meetings app using user reviews on google play store with sentiment annotation of BERT and hybridization of RNN and LSTM. *Expert Systems with Applications*, *223*, Article 119919.

Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., & Muller, P.-A. (2019). Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*, *33*(4), 917–963.

Jia, R., Chamoun, R., Wallenbring, A., Advand, M., Yu, S., Liu, Y., et al. (2023). A spatio-temporal deep learning model for short-term bike-sharing demand prediction. *Electronic Research Archive*, *31*(2), 1031–1047.

Kim, K. (2018). Investigation on the effects of weather and calendar events on bike-sharing according to the trip patterns of bike rentals of stations. *Journal of Transport Geography*, *66*, 309–320.

Kitamura, R., Chen, C., Pendyala, R. M., & Narayanan, R. (2000). Micro-simulation of daily activity-travel patterns for travel demand forecasting. *Transportation*, *27*, 25–51.

Kleinberg, J. (2007). The wireless epidemic. *Nature*, *449*, 287–288.

Kraemer, M., Yang, C., Gutierrez, B., Wu, C., Klein, B., Pigott, D., et al. (2020). The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*, *368*(6490), 493–497. http://dx.doi.org/10.1126/science.abb4218, 2020 May 1.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90.

Li, X., Xu, Y., Chen, Q., Wang, L., Zhang, X., & Shi, W. (2021). Short-term forecast of bicycle usage in bike sharing systems: a spatial-temporal memory network. *IEEE Transactions on Intelligent Transportation Systems*.

Li, X., Xu, Y., Zhang, X., Shi, W., Yue, Y., & Li, Q. (2023). Improving short-term bike sharing demand forecast through an irregular convolutional neural network. *Transportation Research Part C (Emerging Technologies)*, *147*, Article 103984.

Li, D., Zhao, Y., & Li, Y. (2019). Time-series representation and clustering approaches for sharing bike usage mining. *IEEE Access*, *7*, 177856–177863.

Li, Y., Zheng, Y., Zhang, H., & Chen, L. (2015). Traffic prediction in a bike-sharing system. In *Proceedings of the 23rd SIGSPATIAL international conference on advances in geographic information systems* (pp. 1–10).

Li, Y., Zhu, Z., Kong, D., Xu, M., & Zhao, Y. (2019). Learning heterogeneous spatial-temporal representation for bike-sharing demand prediction. In *Proceedings of the AAAI conference on artificial intelligence, vol. 33* (pp. 1004–1011).

Lillicrap, T. P., & Santoro, A. (2019). Backpropagation through time and the brain. *Current Opinion in Neurobiology*, *55*, 82–89.

Lin, L., He, Z., & Peeta, S. (2018). Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach. *Transportation Research Part C (Emerging Technologies)*, *97*, 258–276.

Lu, E. H.-C., & Lin, Z.-Q. (2020). Rental prediction in bicycle-sharing system using recurrent neural network. *IEEE Access*, *8*, 92262–92274.

Lu, W., & Wei, X. (2017). Spatio-temporal distribution pattern of cable car passenger flow in panholidays: A case study of Huangshan Scenic Area. In *2017 IEEE second international conference on data science in cyberspace (DSC)* (pp. 35–42). IEEE.

Luo, X., Wu, H., Wang, Z., Wang, J., & Meng, D. (2021). A novel approach to large-scale dynamically weighted directed network representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(12), 9756–9773.

Lv, Y., Zhi, D., Sun, H., & Qi, G. (2021). Mobility pattern recognition based prediction for the subway station related bike-sharing trips. *Transportation Research Part C (Emerging Technologies)*, *133*, Article 103404.

Márquez, H. R. G., Bracho, R. L., & Ramirez-Nafarrate, A. (2021). A simulation-optimization study of the inventory of a bike-sharing system: The case of Mexico City Ecobici's system. *Case Studies on Transport Policy*, *9*(3), 1059–1072.

Miskolczi, M., Földes, D., Munkácsy, A., & Jászberényi, M. (2021). Urban mobility scenarios until the 2030s. *Sustainable Cities and Society*, *72*, Article 103029.

Noble, W. S. (2006). What is a support vector machine? *Nature biotechnology*, *24*(12), 1565–1567.

Ostertagová, E. (2012). Modelling using polynomial regression. *Procedia Engineering*, *48*, 500–506.

Palangi, H., Deng, L., Shen, Y., Gao, J., He, X., Chen, J., et al. (2016). Deep sentence embedding using long short-term memory networks: Analysis and application to information retrieval. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *24*(4), 694–707.

Pappalardo, L., Simini, F., Rinzivillo, S., Pedreschi, D., Giannotti, F., & Barabási, A.-L. (2015). Returners and explorers dichotomy in human mobility. *Nature Communications*, *6*, 8166. http://dx.doi.org/10.1038/ncomms9166.

Peláez-Rodríguez, C., Pérez-Aracil, J., de Lopez-Diz, A., Casanova-Mateo, C., Fister, D., Jiménez-Fernández, S., et al. (2023). Deep learning ensembles for accurate fog-related low-visibility events forecasting. *Neurocomputing*, *549*, Article 126435.

Peng, Y., Liang, T., Hao, X., Chen, Y., Li, S., & Yi, Y. (2021). CNN-GRU-AM for shared bicycles demand forecasting. *Computational Intelligence and Neuroscience*, *2021*.

Peng, L., Wang, L., Xia, D., & Gao, Q. (2022). Effective energy consumption forecasting using empirical wavelet transform and long short-term memory. *Energy*, *238*, Article 121756.

Rawassizadeh, R., Momeni, E., Dobbins, C., Gharibshah, J., & Pazzani, M. (2016). Scalable daily human behavioral pattern mining from multivariate temporal data. *IEEE Transactions on Knowledge and Data Engineering*, *28*, 11. http://dx.doi.org/10.1109/TKDE.2016.2592527.

Reid, S., Tibshirani, R., & Friedman, J. (2016). A study of error variance estimation in lasso regression. *Statistica Sinica*, 35–67.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536. http://dx.doi.org/10.1038/323533a0.

Salcedo-Sanz, S., García-Herrera, R., Camacho-Gómez, C., Alexandre, E., Carro-Calvo, L., & Jaume-Santero, F. (2019). Near-optimal selection of representative measuring points for robust temperature field reconstruction with the CRO-SL and analogue methods. *Global and Planetary Change*, *178*, 15–34.

Sardinha, C., Finamore, A. C., & Henriques, R. (2021). Context-aware demand prediction in bike sharing systems: Incorporating spatial, meteorological and calendrical context. arXiv preprint arXiv:2105.01125.

Sathishkumar, V., Park, J., & Cho, Y. (2020). Using data mining techniques for bike sharing demand prediction in metropolitan city. *Computer Communications*, *153*, 353–366.

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, *14*(3), 199–222.

Song, C., Qu, Z., Blumm, N., & Barabási, A.-L. (2010). Limits of predictability in human mobility. *Science*, *327*(5968), 1018–1021. http://dx.doi.org/10.1126/science.1177170, 2010 Feb 19.

Souto, G., & Liebig, T. (2016). *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*: *vol. 9580*, *On event detection from spatial time series for urban traffic applications* (pp. 221–233). http://dx.doi.org/10.1007/978-3-319-41706-6_11.

Sun, C., & Lu, J. (2023). The relative roles of different land-use types in bike-sharing demand: A machine learning-based multiple interpolation fusion method. *Information Fusion*.

Sutskever, I., Martens, J., & Hinton, G. E. (2011). Generating text with recurrent neural networks. In *ICML* (pp. 1–8).

Tan, S., Lai, S., Fang, F., Cao, Z., Sai, B., Song, B., et al. (2021). Mobility in China, 2020: a tale of four phases. *National Science Review*, *8*(11), Article nwab148.

Tekouabou, S. C. K., et al. (2021). Intelligent management of bike sharing in smart cities using machine learning and Internet of Things. *Sustainable Cities and Society*, *67*, Article 102702.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *58*(1), 267–288.

Toch, E., Lerner, B., Ben-Zion, E., & Ben-Gal, I. (2019). Analyzing large-scale human mobility data: a survey of machine learning methods and applications. *Knowledge and Information Systems*, *58*(3), 501–523. http://dx.doi.org/10.1007/S10115-018-1186-X/FIGURES/5.

Tong, S., & Koller, D. (2001). Support vector machine active learning with applications to text classification. *Journal of Machine Learning Research*, *2*(Nov), 45–66.

Torres, J. F., Hadjout, D., Sebaa, A., Martínez-Álvarez, F., & Troncoso, A. (2021). Deep learning for time series forecasting: a survey. *Big Data*, *9*(1), 3–21.

Torres-López, R., Casillas-Pérez, D., Pérez-Aracil, J., Cornejo-Bueno, L., Alexandre, E., & Salcedo-Sanz, S. (2022). Analysis of machine learning approaches' performance in prediction problems with human activity patterns. *Mathematics*, *10*(13), 2187.

Trafalis, T. B., & Ince, H. (2000). Support vector machine for regression and applications to financial forecasting. In *Proceedings of the IEEE-INNS-ENNS international joint conference on neural networks. IJCNN 2000. Neural computing: New challenges and perspectives for the new millennium, vol. 6* (pp. 348–353). IEEE.

Van Arem, B., Kirby, H. R., Van Der Vlist, M. J., & Whittaker, J. C. (1997). Recent advances and applications in the field of short-term traffic forecasting. *International Journal of Forecasting*, *13*(1), 1–12.

Vanvyve, E., Delle Monache, L., Monaghan, A. J., & Pinto, J. O. (2015). Wind resource estimates with an analog ensemble approach. *Renewable Energy*, *74*, 761–773.

Ve, S., & Cho, Y. (2020). Season wise bike sharing demand analysis using random forest algorithm. *Computational Intelligence*.

Wang, B., & Kim, I. (2018). Short-term prediction for bike-sharing service using machine learning. *Transportation Research Procedia*, *34*, 171–178.

Wang, D., Pei, X., Li, L., & Yao, D. (2018). Risky driver recognition based on vehicle speed time series. *IEEE Transactions on Human-Machine Systems*, *48*(1), 63–71. http://dx.doi.org/10.1109/THMS.2017.2776605.

Xie, N., Li, Z., Liu, Z., & Tan, S. (2023). A censored semi-bandit model for resource allocation in bike sharing systems. *Expert Systems with Applications*, *216*, Article 119447.

Xu, M., Di, Y., Yang, H., Chen, X., & Zhu, Z. (2023). Multi-task supply-demand prediction and reliability analysis for docked bike-sharing systems via transformer-encoder-based neural processes. *Transportation Research Part C (Emerging Technologies)*, *147*, Article 104015.

Yang, Y., Heppenstall, A., Turner, A., & Comber, A. (2020). Using graph structural information about flows to enhance short-term demand prediction in bike-sharing systems. *Computers, Environment and Urban Systems*, *83*, Article 101521.

Yang, Y., Shao, X., Zhu, Y., Yao, E., Liu, D., Zhao, F., et al. (2023). Short-term forecasting of dockless bike-sharing demand with the built environment and weather. *Journal of Advanced Transportation*, *2023*.

Yang, H., Xie, K., Ozbay, K., Ma, Y., & Wang, Z. (2018). Use of deep learning to predict daily usage of bike sharing systems. *Transportation Research Record*, *2672*(36), 92–102.

Zhang, H., Zhuge, C., Jia, J., Shi, B., & Wang, W. (2021). Green travel mobility of dockless bike-sharing based on trip data in big cities: a spatial network analysis. *Journal of Cleaner Production*, *313*, Article 127930.