

Received August 9, 2021, accepted August 26, 2021, date of publication September 7, 2021, date of current version September 15, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3110794

# Deep Learning for Short-Term Prediction of Available Bikes on Bike-Sharing Stations

ENRICO COLLINI<sup>1</sup>, PAOLO NESI<sup>1</sup>, (Member, IEEE), AND GIANNI PANTALEO<sup>1</sup>

Distributed Systems and Internet Technologies Laboratory, Department of Information Engineering, University of Florence, 50139 Firenze, Italy

Corresponding author: Paolo Nesi (paolo.nesi@unifi.it)

**ABSTRACT** Bike-sharing is adopted as a valid option replacing traditional public transports since they are eco-friendly, prevent traffic congestions, reduce any possible risk of social contacts which happen mostly on public means. However, some problems may occur such as the irregular distribution of bikes on related stations/racks/areas, and the difficulty of knowing in advance what the rack status will be like, or predicting if there will be bikes available in a specific bike-station at a certain time of the day, or if there will be a free slot to leave the rented bike. Thus, providing predictions can be useful to improve the service quality, especially in those cases where bike racks are used for e-bikes, which need to be recharged. This paper compares the state-of-the-art techniques to predict the number of available bikes and free bike-slots in bike-sharing stations (i.e., bike racks). To this end, a set of features and predictive models were compared to identify the best models and predictors for short-term predictions, namely of 15, 30, 45, and 60 minutes. The study has demonstrated that deep learning and in particular Bidirectional Long Short-Term Memory networks (Bi-LSTM) offers a robust approach for the implementation of reliable and fast predictions of available bikes, even with a limited amount of historical data. This paper has also reported an analysis of feature relevance based on SHAP that demonstrated the validity of the model for different cluster behaviours. Both solution and its validation were derived by using data collected in bike-stations in the cities of Siena and Pisa (Italy), in the context of Sii-Mobility National Research Project on Mobility and Transport and Snap4City Smart City IoT infrastructure.

**INDEX TERMS** Available bikes prediction, bike-sharing, deep learning, machine learning, prediction models, smart city.

## I. INTRODUCTION

Cities are becoming large and complex entities. Today, about 55% of the worldwide population lives in urban areas, and that figure is expected to achieve 68% in 2050, according to the “World Urbanization Prospects 2018”, published by the United Nations Department of Economics and Social Affairs [1]. This growth motivates the need to build cities more liveable and sustainable, with modern infrastructures capable to offer smart systems to the citizens [2]. Transportation is one of the most important causes of air pollution, while a more efficient use of bikes may represent a part of the solution. Therefore, bike-sharing systems are widely used in many cities, offering a sustainable alternative and complement to public transport [3], [4], to reduce congestion [5]. Bike stations can detect the presence of bikes with their status, recharge them, indicate when they are ready to

be rented again. Another option could be based on free floating bike-sharing where bikes are more intelligent and can communicate their position to the central management servers, as it occurs with the Mobike solution. On this latter model, e-bikes are more complex to be managed since they are left on the road and the recharge phase has to be implemented by some personnel with a relevant effort almost every day.

In this article, the solution with bikes and rack stations is addressed. Bikes can be typically released at any station whenever a free slot is available. A full station may cause trouble to users discovering that, when arriving to the rack and then forced to look for another bike rack. One of the problems of bike-sharing is related to the irregular distribution of bikes among the various stations and the impossibility of knowing exactly, or at least with a certain degree of probability, if there is bike availability at a desired station in a precise time slot of the day, or just a few minutes in advance. Authors in [6]–[8] used optimization methods to find the best path for operators

The associate editor coordinating the review of this manuscript and approving it for publication was Mehdi Hosseinzadeh<sup>1</sup>.

to rebalance bike stations. On the other hand, in [9] they studied the inventory optimization in a bike-sharing system. A more dynamic approach could be used if operators could know the future status of bike racks 1 *hour* in advance by means of short-term predictions. The same thing applies in favour of users as to the possibility of finding a bike to rent or return in a station with free slots. Therefore, predicting bike availability (as well as free slots) per station over time can be useful to manage the demands for bikes per station, and perform the redistribution in advance.

In literature, urban bike-sharing systems have attracted extensive research efforts in past decades and nowadays they still are a growingly active research topic. In [10], a mathematical model has been proposed to determine the number of needed docking stations, their locations and the possible cycle path network, as well as models to make predictions about possible routes taken by users between stations of origin and destination. In [11] and in [12], clustering and forecasting techniques have been used on the network of bike-sharing stations in Barcelona to obtain useful information to describe the city mobility. In [13], authors interpreted the system as a dynamic network by analysing how bicycle flows distribute spatially along the network. In [14], different bike-sharing services are analysed to highlight the differences in bike flows and routes. In [15], clustering techniques have been used to study the docking network stations in Wien. The above presented cases studied both dynamics and behaviour of bike-sharing systems. This study concerns a specific area of interest: the prediction of bike-sharing related metrics such as, the number of bikes available in bike sharing systems with smart stations, the number of check-in and check-out in stations as in [16], or the prediction of demand within a station as in [17], and [18] how to carry out preventive measures.

## A. ARTICLE CONTRIBUTIONS AND STRUCTURE

The main contribution of this paper consists in presenting a solution which compares the state-of-the-art technologies for short-term prediction (15, 30, 45, and 60 minutes) as to available bikes on bike-sharing stations, and thus the number of free slots according to the size of the station and the number of broken bikes, within the cities of Siena and Pisa in Italy and considering a limited data history of 3 months.

Prediction of available bikes is a non-linear process whose dynamic changes involve multiple kinds of factors, resulting from the context. To this end, the solution has been based on different cities and locations, and despite the changes in Siena and Pisa, the same model has been used and the same features have been identified in both cases, thus demonstrating the validity of the obtained outcomes. The validation has been performed also using XAI approach (explainable artificial intelligence) to understand the relevance of features and how this could change in different clusters. The analysis has proven both validity and flexibility of the model which has obtained good results for different clusters and thus for different behaviours of time series, according to the identified features and different relevance patterns.

The solutions have been implemented in the context of the Sii-Mobility<sup>1</sup> project and infrastructure (national mobility and transport smart city project of Italian Ministry of Research for terrestrial mobility and transport), which is a solution based on the Km4City<sup>2</sup> model and Snap4City<sup>3</sup> tools [19]–[21]. Sii-Mobility is currently covering the whole Tuscany region (Italy), which hosts 3.5 inhabitants and 40 million tourists per year. The Sii-Mobility project aimed at defining solutions for sustainable mobility, suggesting bikes availability status to users at least 15 minutes - 1 hour in advance to allow them to take a conscious decision, and maybe change their own travel schedule. The focus reported on the paper is related to bike-sharing services in the cities of Siena and Pisa within the Tuscany region of Italy.

The paper is structured as follows. In Section II, the related works are presented and discussed. Section III provides a description of the bike-sharing data, and their characterization in terms of group clustering. In addition, the identification of several features being the basis of the suggested predictive models is reported. In Section IV, there are the machine learning approaches adopted to identify and validate the predictive models and framework. Conclusions are drawn in Section V.

## II. RELATED WORKS

The problem of Bike-Sharing related metrics such as bike availability or bike usage etc. has been addressed through different approaches. The most recent works on the state-of-the-art use machine learning methodologies, in particular ensemble learning techniques such as Random Forest (RF), Gradient Boosting Machines (GB) and Deep learning methods such as Deep Neural Network (DNN). On the other hand, on other prediction problems regarding exploiting time-series, the most used deep learning approaches use deep recurrent neural networks because of their capability of using not only the information at any precise instant, but also pieces of information from previous steps. In **Table 1**, a summary of the state-of-the-art solutions, when considering bike-sharing systems with stations, is reported.

In [22], DNNs have been used to predict the number of riders for a bike rental company with a prediction target of 1 hour. The input data included also meteorological features as temperature, humidity, and wind speed, and data derived from time information such as season, year, month, hour. They stated that deeper architectures lead to better results from 70% of accuracy with only 1 hidden layer up to 80% with deeper neural networks.

In [23], the author compared ensemble learning techniques Extreme GB (XGBoost) Regression Tree (XGBoost tree) and RF with a DNN for the hourly prediction of the number of bike changes in the stations of City Bike in North America. Two methods were proposed to calculate the number of bike

<sup>1</sup><https://www.sii-mobility.org>

<sup>2</sup><https://www.km4city.org>

<sup>3</sup><https://www.snap4city.org>

**TABLE 1. Comparison of related work solutions, with main attention to deep learning aspects and better results.**

citation	Target	Features	Dataset	Model	Reported Best Results		
					RF	Rentals	returns
[24]	1h, 2h, 3h bike rentals and returns	Bike rented, Bike returned, Avg temperature, Wind speed, Sky cover, Rain, holiday or Sunday, time, weekday, month, year	ThessBike	RF, XGBoost, GB, DNN	MAE 0.85	0.82	0.82
[23]	Hourly Bike number change in station	Usage features, spatial features, temporal features	Citi Bike dataset July – August 2017	XGBoost tree, RF, DNN	XGBoost tree MAE 1.8159 AP 0.7085		
[25]	1h rental bikes rented	Rental bikes rented, Weekend/weekday, Day of the week, Holidays, Functional/non functional, Temperature, Humidity, Windspeed, Visibility, Dew Point, temperature, Rainfall, snowfall	Seoul (South Korea)	RF, SVM, k-Nearest neighbours (KNN), Classification and Regression Trees (CART)	RF results: R2 0.88 RMSE 216.01 MAE 130.52 CV 30.63 PI 0.73		
[26]	Hourly rental bike demand	Temperature, Humidity, Windspeed, Visibility, Dewpoint, Solar radiation, Snowfall, Rainfall, number of bikes rented per hour, date information.	Seoul (South Korea)	LR, XGBoost, SVM, Boosted Trees, XGBoost Trees	XGBoost results: R2 0.92 RMSE 174.68 MAE 109.89 CV 24.92		
[27]	Long terms predictions	Timestamp, count of new bike shared, temperature, humidity, windspeed, weather code, is holiday, is weekend, season	London	LR, RF, XGBoost, SVM, AB, BGR	RF results: R2 0.95 MAE 0.04 MSE 0.01 RMSLE 0.03		
[22]	1h number of rider	Number of riders, Season, year, month, hour, day, holiday, weekday, working, weather, Temperature, humidity, wind speed	Rental Company unspecified	DNN	80% accuracy		
[28]	1,5, 10 minutes number of bikes in the station	Station ID, Number of available bikes, timestamp	Suzhou	RF, LSTM, GRU	Graphical results reported		
[29]	Next hour rental and return demand	Year, Month, Weekday, Hour, Quarter-hour, Hourly weather status, Hourly temperature, humidity, intensity of precipitation, prob of precipitation, dew point, wind speed, visibility, Number of check in bikes from subscribers and from customers, Number of checkout bikes from subscribers and from customers	New York City Bike Dataset	FFNN + BI-RNN (LSTM or GRU) + soft Attention mechanism	GRU MAE 1.11 RMSE 2.06 RMSLE 0.46 Global MAPE 0.14 GlobalRMSLE 0.21		
[30]	10 minutes bikes available	Station name, Station location, Time, total rack count, bikes available	Daejeon, Republic of Korea	RF, XGBoost, LSTM, GRU	GRU results: MSE 0.002336 MAE 0.24275		

changes in stations. The first one generates the prediction of Check-out and Check-in bikes and through the difference calculates the target. The second one uses the number of bike change in station as a feature for the predictive model. Training the models with the second approach leads to better results and in particular the best algorithm was the XGBoost Regressor.

In [25], 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. They found that the trend of the bike demand is similar in Spring and Autumn, while it assumes a different pattern in Winter and Summer. Developing specific season-wise predictive models improves the results if compared to an overall model.

In [26], researchers investigated which regression model is the best for the hourly bike sharing demand prediction among Linear Regression (LR), GB, Support Vector Machine (SVM), Boosted Trees and XGBoost for the city of Seoul, South Korea. The data used included meteorological features (Temperature, Humidity, Windspeed, Visibility, Dewpoint, Solar radiation, Snowfall, Rainfall), the number of bikes rented per hour and date information. For each model a filtering of the least important features has been applied

so as to eliminate unproductive parameters and for each regression model. The hyperparameters for every model have been optimized. The XGBoost regressor achieved the best results in terms of R-squared ( $R^2$ ) Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and coefficient of variance (CV). Researchers reported that Temperature and Hour are the most significant variables for the hourly rental bike count prediction in each and every model.

In [27], researchers set the problem of predicting the number of bikes shared per hour, day and month in London with machine learning regressors techniques. The exploited algorithms are RF, Bagging regressor (BGR), XGBoost, and Ada Boosting (AB) regressor. The set of features included the data related to the bike-sharing system, meteorological measures, and time information like if the day under observation was a working day, a holiday or a weekend. RF, BGR and XGBoost achieved the best performance in terms of R2, MAE, Mean Squared Error (MSE) and Root Mean Squared Log Error (RMSLE).

In [24], learning techniques RF and XGBoost achieved better results compared to the DNN implemented for the target of predicting rented and returned bikes for the next 1,2, and 3 hours in Thessaloniki, Greece. The predictions of such developed models have been incorporated in a tool that can be used by operators of the bike system.

Due to the network structure of bike-sharing systems [42] a Graph Convolutional Neural Network (GCNN) has been used in order to predict the station-level hourly demand in New York. The designed architecture is able to capture not only the temporal dependencies in bike-sharing demand series, but also the heterogeneous pairwise correlations among stations.

Deep learning techniques have been often used for the prediction of problems on time-series. The state-of-the-art solutions regard neural networks with a recurrent architecture. This type of networks is best suited for this type of problems for its capability of using the information at the current instant, as well as the information from previous steps.

In [28], RF, Long Short-Term Memory networks (LSTM) and Gated Recurrent Units networks (GRU) have been compared for the short-term prediction of the number of available bikes in different time ranges of 1, 5 and 10 minutes, in Suzhou, China. The dataset included only features related to the bike-sharing system. The Recurrent Neural Networks, RNNs, performed better than the ensemble learning techniques.

The same conclusions have been also reported for the bike-sharing system in Daejeon, Republic of Korea [30]. Researchers used 261 datasets of Tashu Bike stations and tested the ensemble learning techniques RF and XGBoost with the deep learning recurrent neural networks: LSTM and GRU. The features used for the dataset are Station name, Station location, Time, total rack size and available bikes. The GRU model had the lowest error predictions about average MSE and MAE.

In [29], Bidirectional Recurrent Neural Network (BI-RNN) architecture has been used for predicting rental and return demand in the forthcoming hour on New York City Bike Dataset. The features used included data of the bike-sharing systems, data information and meteorological metrics. The BI-RNN has been compared Ordinary Least-Squares Regression (OLSR), RF, and a Feedforward Neural Network, FFNN, with 4 layers and it achieved the best results on the test set in terms of MAE, RMSE, RMSLE, MAPE.

### III. CLUSTERING, DATA DESCRIPTION

As mentioned in the introduction, the main goal was to find a solution to predict bike availability for the bike stations of Siena and Pisa using only a limited history of data in the Tuscany region in Italy. The total number of bike-racks is 39, 24 of which are in Pisa and 15 in Siena. The status of each station is registered every 15 minutes including the total capacity of the rack, the number of broken bikes and the number of available bikes. The data refer to 15 stations located in the municipality of Siena and 24 located in Pisa, from 2019 to March 2020, see Section IV.A for more details.

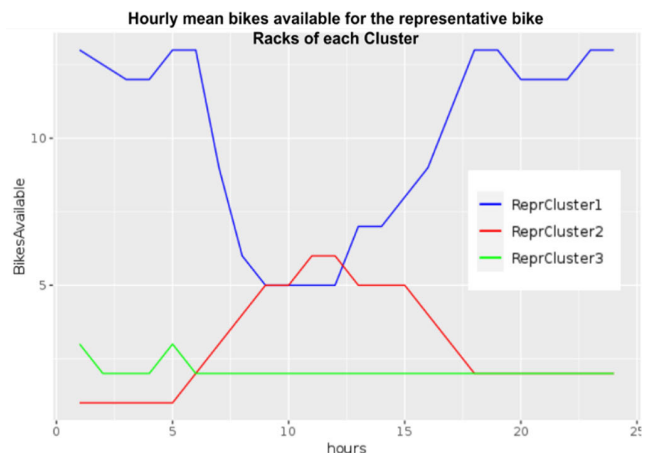
As a first step, a clustering approach has been applied in order to classify bike stations on the basis of their bike availability trend. The aim of the clustering was to identify the number of typical trends over time and test the prediction models for those cases. The K-means clustering method [31]

has been applied to identify clusters based on the mean hourly trend of the number of available bikes (normalized between 0 and 1) in the considered history. The optimal number of clusters resulted to be equal to 3 and it was identified by using the Elbow criteria [32]. Each cluster represents a group of stations located in a particular area of Siena/Pisa municipalities. Clusters include bike racks of both cities and they characterize the trend only. In Fig. 1, the hourly trends of the representative sensors related to the computed clusters are reported. In Fig. 2 and Fig. 3 the position of bike racks on the map for Siena and Pisa is depicted. The number on each marker on the map is related to the identifier of the bike rack and the font color is white for all the representative racks.

Please note that stations/racks belonging to

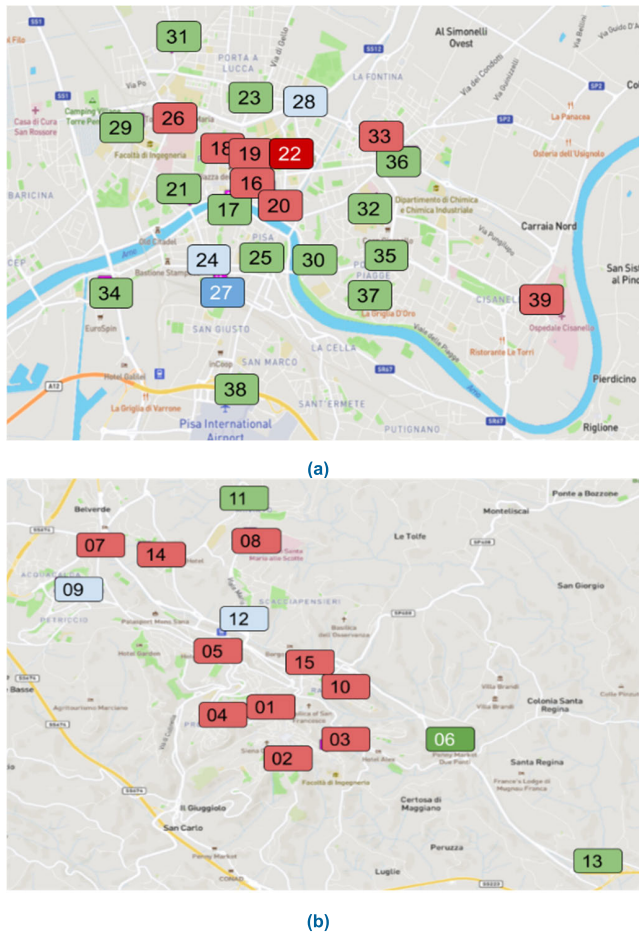
- **Cluster 1** are typically characterized by a decrement of bike availability at lunchtime and they are mainly located close to the railway stations, airport, etc. The representative bike rack is the Bike rack *Stazione F.S.* of Pisa.
- **Cluster 2** are typically positioned in the central area of the cities and they are characterized by an increment of the availability of bikes in the central part of the day (lunch hours, since most people are parking their bikes to get lunch). The representative bike rack is the Bike rack *Polo Marzotto* of Pisa.
- **Cluster 3** presents an almost uniform trend in the bike availability and bike racks are mainly positioned in the peripheral areas of the city. The representative rack is the Bike rack *Due Ponti* in Siena.

For example, in Siena municipality “*Terminal Bus*” (id = 12) is a bike-sharing station positioned nearby the train station in Siena and it actually belongs to Cluster 1. In Pisa, the “*Duomo*” station (id = 26) belongs to Cluster 2 and it is actually positioned in the city centre, while “*Pietrasantina*” (id = 31) belongs to the third cluster and it is indeed located in a peripheral area of Pisa.



**FIGURE 1.** Hourly bikes available for the representative rack of the clusters determined by the K-means. The trend of the representative bike rack for Cluster 1 is displayed in blue, the one for Cluster 2 in red, while green is for Cluster 3.



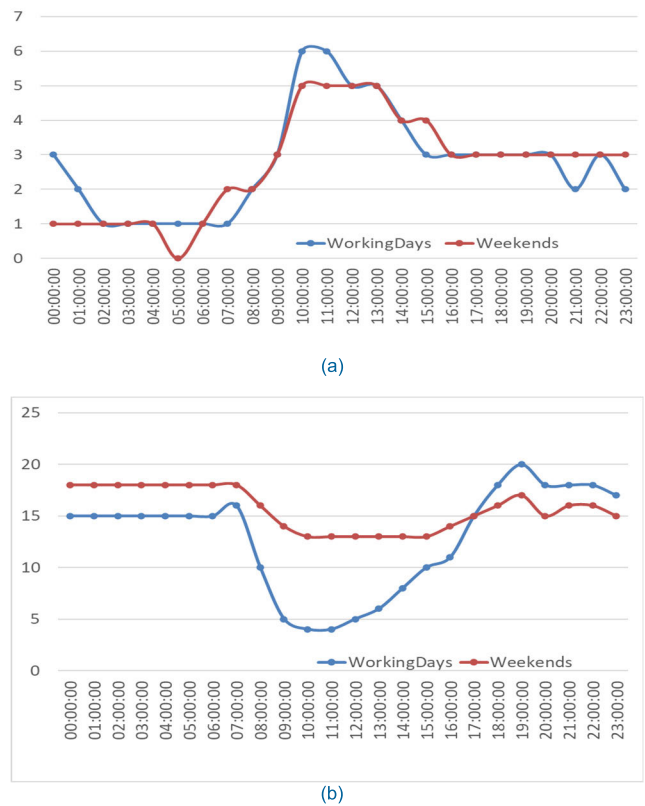


**FIGURE 2.** Positioning of bike racks in Pisa (a) and in Siena (b). The colour depends on the specific cluster: blue for Cluster 1, red for Cluster 2 and green for Cluster 3. If the selected rack is representative of its cluster, the font colour is white and the background is darker.

Moreover, we have also detected some changes in the typical trends between working days and weekends, as shown in Fig. 3. Fig. 3 (a) focusses on the comparison between trends for working days and weekends as to “Curtatone” station in Siena, while Fig. 3 (b) shows trends of working days/weekends as to “Stazione F.S” in Pisa.

With the aim of developing a predictive model, a set of features has been identified and tested (see Table 2). Features belonging to the **Baseline (time series)** category refer to aspects related to the direct observation of bike rack status over time, for instance: date and time when measurements are taken, information whether it is a working day or not, number of bikes on racks, etc. Typically, these values are recorded every 15 minutes. Features describing the **differences over time**. Usually, as to the number of bikes the trend is similar for the same day of the week, as well as for the same day of the month. Thus, other features have been included in the model. Considering  $d$  the observation day in the time slot  $t$ , the included features are as follows.

- $dP$ : the difference between the number of available bikes in the observation day ( $d$ ) at time  $t$  and the number of



**FIGURE 3.** Number of available bike in working days/weekend trends of the (a) “Curtatone” bike-sharing stations in Siena and (b) “Stazione F.S” stations in Pisa.

available bikes during the previous time slot ( $t - 1$ ) of the previous day ( $d - 1$ ).

$$dP = availableBikes_{d,t} - availableBikes_{d-1,t-1}$$

- $dS$  the difference between the number of available bikes in the observation day ( $d$ ) at the time slot  $t$  and the number of bikes during the successive time slot ( $t + 1$ ) of the previous day ( $d - 1$ ).

$$dS = availableBikes_{d,t} - availableBikes_{d-1,t+1}$$

- $PwAB$  the number of available bikes of the previous week ( $d - 7$ ) in the same time slot ( $t$ ).

$$PwAB = availableBikes_{d,t} - availableBikes_{d-7,t}$$

**Real-time weather and weather forecasts** were also collected every 15 minutes (i.e., temperature, humidity and rainfall). Weather information can indeed improve the performance of predictive models as shown in research papers [33] and [34]. It is worth noting that, according to our analysis, significant weather values are the ones related to the current time and the hour just before the measured bike availability time. For example, in order to predict the number of available bikes at a bike rack at 3:00 pm, the weather features at 2:00 pm and at the current time are relevant. In fact, weather conditions typically have an influence on the user’s decision as to either ride a bike or take other means of transportation.

In addition, the weather forecast is relevant, too, for people may plan to use the bike according to the weather forecast of that day.

**IV. PREDICTION MODELS**

In this section, the considered machine learning techniques are compared, with the aim of creating a solution to predict the number of available bikes for the representative bike-racks (resulted from the above presented clustering process) with temporal targets of 15, 30, 45 and 60 minutes. Ensemble learning techniques such as RF [35] and

**TABLE 2. Overview of the features used in the short-term prediction models.**

Category	Feature	Description
target	AB	Number of available Bikes
Baseline-Historical	Time	The observation time hh-mm-ss
	month	Month of observation {1-12}
	Day Of The Week	Day of the week {1-7}
	Weekend	1 if the observation day is Saturday of Sunday, 0 otherwise
Differences Over Time	dP	the difference between the number of available bikes in the observation day ( <i>d</i> ) at the time slot <i>t</i> and the number of available bikes during the previous time slot ( <i>t-1</i> ) of the previous day ( <i>d-1</i> )
	dS	the difference between the number of available bikes in the observation day ( <i>d</i> ) at the time slot <i>t</i> and the number of bikes during the successive time slot ( <i>t+1</i> ) of the previous day ( <i>d-1</i> ).
	PwAB	the number of available bikes of the previous week ( <i>d-7</i> ) in the same time slot ( <i>t</i> ).
Real-time weather and weather forecast	Temperature	Air temperature at the observation time, in °C
	Max Temperature	Forecast of max temperature of the observation day, in °C
	Min Temperature	Forecast of Min temperature of the observation day, in °C
	Humidity	Humidity of the hour prior to the observation time, in percentage
	Rain	mm of rain registered in the hour prior to the observation time
	Pressure	Pressure at the observation time, in millibar (mb)
	Wind Speed	Average wind speed registered in the hour prior to the observation time, in km/h
Cloud Cover Percentage	Cloud Cover Percentage at the observation time	

XGBoost [36] are powerful techniques that should be considered for this type of problem. As to the deep learning techniques for this work, we have compared the DNN architecture with LSTM [38], [39] and based on the results of the related works also with a Deep Bidirectional-LSTM (Bi-LSTM) Neural Network [40].

These models were evaluated in terms of statistical measures such as R-squared (R<sup>2</sup>), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). This last metric is the one that has been used to compare these techniques and choose the best model architecture for the task of short-term predictions of available bikes.

The R<sup>2</sup> is calculated as follows:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n obs_i$$

$$R^2 = 1 - \left( \frac{\sum_{i=1}^n (obs_i - pred_i)^2}{\sum_{i=1}^n (obs_i - \bar{y})^2} \right)$$

The MAE is calculated as follows:

$$MAE = \frac{\sum_{i=1}^n |obs_i - pred_i|}{n}$$

The MAPE is calculated as follows:

$$MAPE = \frac{\sum_{i=1}^n \left| \frac{obs_i - pred_i}{obs_i} \right|}{n} * 100$$

The RMSE is calculated as the Root square of the Mean Squared Error (MSE):

$$MSE = \frac{\sum_{i=1}^n (obs_i - pred_i)^2}{n}$$

$$RMSE = \sqrt{MSE}$$

where:

- obs<sub>i</sub>* = observation at time *i*,
- pred<sub>t</sub>* = prediction at time *t*,
- n* is the number of the values in the test set.

Regarding the implementation of the ensemble learning techniques, the number of trees parameter for the RF was set to 300, with a minimum sample split set equal to 2, minimum

**TABLE 3. Comparison of resulting MAPE obtained by using long and short range periods of training, best results in bold, C1: Cluster 1, C2: Cluster 2, C3: Cluster 3.**

Technique	Results obtained for 14-18 months training set			Results obtained for 3 months training set		
	C1	C2	C3	C1	C2	C3
RF	53.16	<b>142.07</b>	<b>34.08</b>	59.57	238.49	36.49
XGBoost	<b>48.26</b>	<b>107.21</b>	<b>28.47</b>	49.09	119.56	29.36
DNN	<b>46.71</b>	210.87	30.73	49.56	<b>178.23</b>	<b>28.68</b>
LSTM	53.00	232.89	29.86	<b>51.16</b>	<b>164.00</b>	<b>25.88</b>
Bi-LSTM	58.30	144.46	32.22	<b>45.53</b>	<b>92.62</b>	<b>27.32</b>

number of samples allowed for a leaf equal to 1, without limits on the maximum number of features considered to split a node as well as on the number of leaves, with the construction of bootstrapped datasets to create the related trees.

The XGBoost regressor uses the least-squares loss function with learning rate optimized with values 0.1, 0.01, and 0.001 with max depth equal to 3 and minimum sample split, minimum sample leaf, maximum number of features equal to the ones chosen for the RF.

The architecture of the Deep Learning neural networks is made up of 4 layers with specific units of the selected architecture (e.g.: LSTM units for LSTM networks) and optimized hyperparameters. A hyperparameter optimization procedure, based on randomized search, was performed to obtain the best model in terms of predictive performances, considering different parameters, as described hereafter. The number of neurons for the input layer is equal to 64 or 128; for the 2<sup>nd</sup> layer 64, 32; for the 3<sup>rd</sup> layer 16, 32. The last layer has only one neuron with a sigmoid activation function, in order to obtain a value in the range 0, 1 (the input data for the models were normalized using a Min Max scaler). The batch size was set to 32 and 64 samples. Also, the dropout rate for each layer was optimized with the values 0.1, 0.25, 0.5. For each model, the Adam Optimizer has been chosen with learning rate optimized among 0.05, 0.005, 0.0005 and 0.00005. MSE was selected as loss function to be monitored during the optimization. The number of epochs was set to a maximum value of 1000, because the training strategy used the Early Stopping method for determining the optimum epoch number minimizing the RMSE of the validation set, restoring the weights of the best model at the end of the learning process. As to LSTMs and Bi-LSTMs inputs were organized through a sliding window with 4 timesteps, which is equivalent to the values of the previous hour with respect to the prediction time.

#### A. EXPERIMENTAL RESULTS

The data used for this training range from the 16<sup>th</sup> of December 2019 to the 9<sup>th</sup> of February 2020. The successive two weeks (10/02/2020 – 23/02/2020) have been used for the validation and the test set includes data from the 24<sup>th</sup> of February 2020 to the 8<sup>th</sup> of March 2020. In reality, we have

had much longer time periods of data into Snap4City.org platform and service. We have initially considered a longer time range for training, 15 months for Pisa bike racks and 18 months for Siena in the 2019-2020. We tried different training sets progressively shorter, so as to identify the best precision. The final choice made, was to use only 3 months of data, collected from a bike-sharing system with smart stations, which could be sufficient to obtain useful predictions on the number of available bikes in a short-term period. This evidence has been reported in **Table 3** where results in terms of MAPE for 60' predictions are comparing different techniques by using training set of 14-18 months and 3 months, according to a test set as above described. As reported in **Table 3**, deep learning methodologies perform better than other machine learning techniques when considering 3 months of training.

On the basis of the short-training data, the machine learning solutions were compared based on the MAPE for the prediction targets of 15, 30, 45 and 60 minutes. Results for the representative bike racks from related clusters are reported in **Table 4**. The minimum MAPE was registered for Cluster 1 for the prediction targets by Bi-LSTM. As to Cluster 2, Bi-LSTM architecture performed better than the others, except for the prediction target of 45 minutes where the unidirectional LSTM network achieved a MAPE of 79.3. As to Cluster 3, Bi-LSTM achieved better results for the prediction target of 15 and 30 minutes, while LSTM achieved the minimum MAPEs for the targets of 30 and 60 minutes. In general, Deep Recurrent Neural Networks architectures outperformed the ensemble learning techniques, despite the limited amount of data and the DNN. Overall, the best machine learning technique for the prediction of the number of available bikes turned out to be the Bi-LSTM. The details on the hyperparameters resulting from Random Search Optimization of Bi-LSTM for the temporal target of 60 minutes are reported in **Table 5**.

The  $R^2$ , RMSE, MAE, MAPE obtained by Bi-LSTM for prediction targets of 15, 30, 45 and 60 minutes are reported in **Figure 4**. Considering the MAE metric, predictions made by Bi-LSTM range between 0.87 (for the prediction target of 15 minutes on Cluster 2) up to 2 bikes (for the temporal target of 60 minutes on Cluster 1 **Fig. 5 (a, b)** reports two

**TABLE 4.** MAPE for the prediction of available bikes in the next 15, 30, 45, 60 minutes on the representative racks of clusters. Comparison of: RF, XGBoost, DNN, LSTM, and Bi-LSTM. In bold the best values of MAPE.

Comparative Results	Cluster1:				Cluster2:				Cluster3:			
	15'	30'	45'	60'	15'	30'	45'	60'	15'	30'	45'	60'
RF	35.16	44.93	53.73	59.57	107.03	146.16	196.55	238.49	30.29	31.60	35.13	36.49
XGBoost	18.75	27.16	40.33	49.09	58.43	83.54	112.46	119.56	28.62	27.30	26.97	29.36
DNN	21.12	28.39	36.01	49.56	109.69	127.23	149.84	178.23	30.29	28.00	27.98	28.68
LSTM	17.68	40.56	44.54	51.16	85.09	120.00	<b>79.30</b>	164.00	22.13	<b>22.91</b>	26.21	<b>25.88</b>
Bi-LSTM	<b>16.46</b>	<b>25.35</b>	<b>33.00</b>	<b>45.53</b>	<b>52.18</b>	<b>63.45</b>	132.00	<b>92.62</b>	<b>21.98</b>	23.00	<b>25.15</b>	27.32

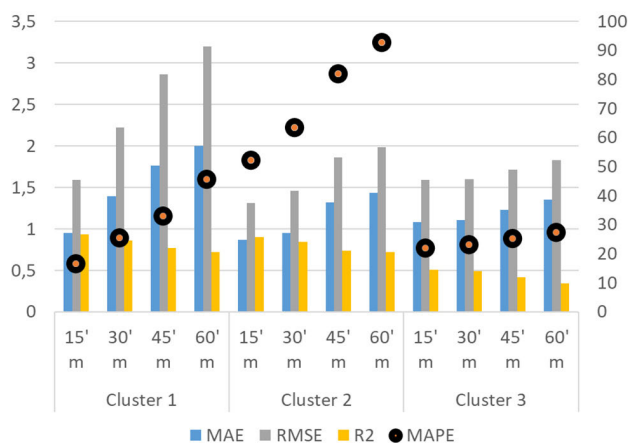
**TABLE 5.** Hyperparameter optimized for the Bidirectional LSTM for the prediction target of 60 minutes.

negMS E	Unit s 1 <sup>st</sup> layer	Unit s 2 <sup>nd</sup> layer	Unit s 3 <sup>rd</sup> layer	Dropou t Rate	Learnin g Rate	Batc h Dim
<b>Cluster 1</b>						
-0.014	64	32	32	0.5	0.0005	32
-0.016	128	32	32	0.1	0.005	64
-0.44	64	64	16	0.25	0.0005	64
<b>Cluster 2</b>						
-0.011	64	64	32	0.1	0.00005	32
-0.012	64	64	16	0.5	0.0005	32
-0.019	64	32	32	0.1	0.05	64
<b>Cluster 3</b>						
-0.013	32	32	16	0.0005	0.5	32
-0.015	64	32	32	0.005	0.25	64
-0.016	64	64	16	0.00005	0.1	64

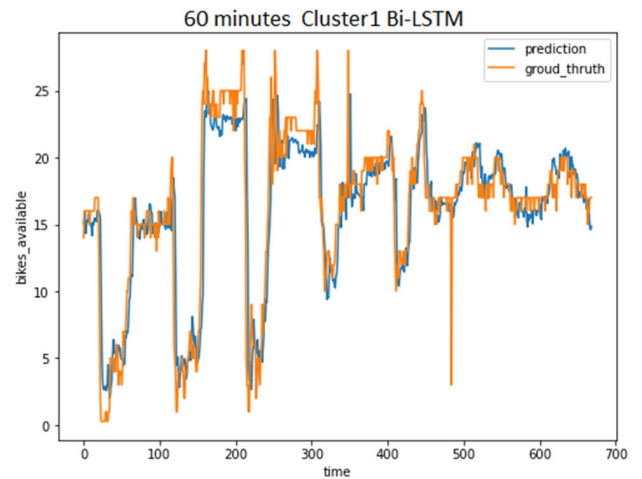
examples about predictions made by Bi-LSTM for temporal target of 60 minutes on the number of available bikes for the representative racks of clusters 1 and 3, respectively.

**B. COMPARISON WITH THE STATE OF THE ART**

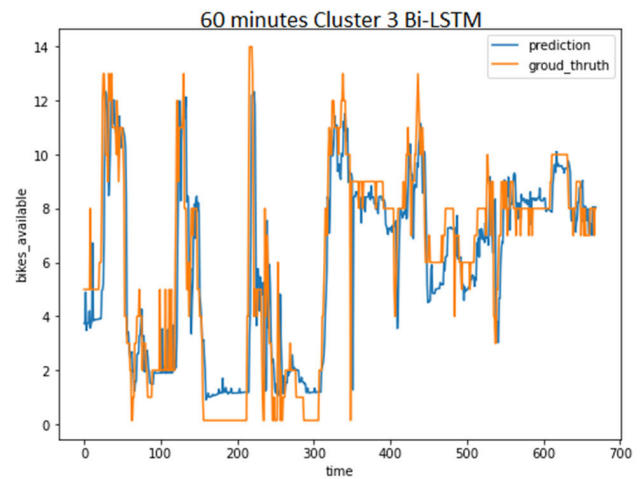
Let us now compare the obtained results which are presented in this paper with respect to the above-presented research papers as to the state-of-the-art for the short-term bike-sharing related metrics prediction. It is evident that the target solution is not a simple task because bike-sharing systems configured with smart stations have many contextual aspects that are case dependent. For example, the number of bikes on the racks can vary among different bike-sharing systems and from station to station. The bike-sharing metrics chosen as prediction targets can vary and are related to the studied case. Temporal targets for bike-sharing related metrics predictions range from 1, 5, 10 minutes up to 1, 2, 3 hours, yet remaining in the context of short-term predictions and the metrics used to evaluate results are different, too. In effect, a reasonable prediction should be in the range of reaching or



**FIGURE 4.** Resulting metrics for the Bidirectional LSTM on the prediction targets of 15, 30, 45, 60 minutes on Clusters. (MAPE values are on the right).



(a)



(b)

**FIGURE 5.** Plots of predictions made by Bi-LSTM as to bike racks for (a) Cluster 1, (b) Cluster 3.

moving to reach the bike rack, probably from 5-15 minutes to 1 hour.

The solutions reported in Table 1 proposed predictions of different metrics, namely: number of bikes rented and returned, number of bike changes within stations, rental bike demand and available bikes. The number of rented bikes is more relevant for providers, while other measures can be somehow reconducted to the number of available bikes, once each bike rack size is known. The research paper [30] predicts the number of available bikes only 10 minutes in advance. The comparison is difficult to be performed, since their results are reported in terms of MAE and MSE that depend on the number of bikes within the system. What is similar is the consideration on models. Indeed, deep learning techniques achieved better results in both papers than ensemble learning techniques.

Papers dealing with the station’s bike demand, namely [26], [27], considered ensemble learning techniques, which have been outperformed by deep learning strategies in the case we studied. Paper [25] as well uses ensemble learning techniques



and as prediction target a cumulated metric, the number of bikes rented in the bike-sharing system, while we predict a punctual value that is relevant for the user and it is a more complex piece of information to be predicted, as it is a disaggregated information per rack. Paper [29] used a bidirectional recurrent neural network to predict the number of check-in/out bikes for the next hour. Assuming that these are bike-sharing areas, their metric is related to the number of available bikes on the rack. The assessment is performed in terms of MAE, which, as to the check-in/out, resulted in an error bigger than 1 bike. In order to provide a global assessment, the MAE is high, while in our case we obtained a similar MAE for each single rack.

C. SHAP FEATURE IMPORTANCE ANALYSIS

To evaluate the relevance of features used by Bi-LSTMs for short-term bike availability prediction on the representative bike racks of Pisa and Siena, a SHapley Additive exPlanations (SHAP) feature importance analysis was performed [41]. Features with corresponding larger absolute Shapley values are the most important ones. The feature importance has been evaluated with respect to the representative sensors of the three clusters taken into account. The resulting Shapley values in descending order are reported in the plots of Fig. 6 for the three representative clusters.

The most important feature is the same for each cluster and it is the number of available bikes at the observation time (and before according to LSTM model). The top 5 most important features for Cluster 1 include *month and weekend* (temporal feature), *temperature* (meteorological feature) and information about the difference among past values concerning the number of available bikes calculated as the difference between the number of available bikes in the observation day ( $d$ ) at time slot  $t$  and the number of bikes during the successive time slot ( $t + 1$ ) of the previous day ( $d - 1$ )  $dS$  and the previous week number of available bikes ( $d - 7$ ) at the same time slot ( $t$ )  $PwAB$ .

The features regarding the difference in the number of available bikes are ranked not so important for cluster 2 and 3 which, instead, in the top 5 include other additional temporal features, the *Day Of The Week* and the *weekend*, and a meteorological feature, the *pressure*.

The difference in value between the first most important feature and the others is quite relevant. So that it has been reasonable to assess the effect of using only univariate data as input for Bi-LSTMs, instead of using multivariate data as above presented. To this end, a new Bi-LSTM model has been trained using only the most relevant feature: the number of available bikes at the observation time and tested for 1-hour prediction for the representative sensors, by means of MAE, MAPE, RMSE, R2 metrics. The comparative results are reported in Fig. 7 and they provide a measure of the impact of using all the features. Such results have demonstrated that including historical and meteorological features did allow Bi-LSTMs to perform better than to use only the number of available bikes in each and every cluster.

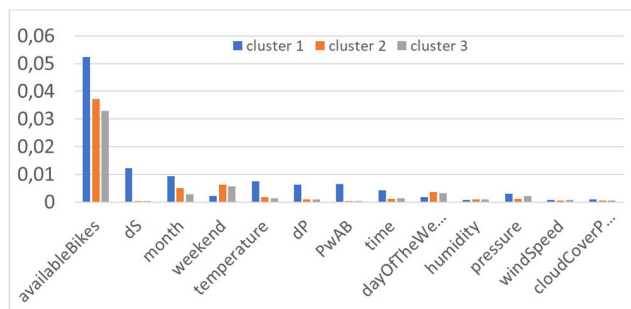


FIGURE 6. Feature Importance graphs. The blue bar plot refers to Cluster 1, the red one to Cluster 2 and the green one to Cluster 3.

D. ECONOMICAL IMPACT

As mentioned in the introduction, most predictive models for bike-sharing are devoted to the prediction of the solution’s global usage such as [25], rather than the prediction of the bike rack status. The prediction of the bike rack status may be used by operators to redistribute bikes during the day instead of waiting for the night (which is called passive rebalancing), as well as to identify potentially all the racks which would need to be adjusted in size. On the other hand, the prediction of available bikes over time is an issue of minor interest for operators, since every time a biker finds a full rack, he/she is constrained to move to an alternative bike rack, thus consuming more minutes and contributing to the redistribution of bikes on racks, which is a double benefit for the operator.

The rebalancing of bike distribution on racks can benefit from both an accurate prediction and user mobility, while taking into account their origin-destination matrices. An active rebalancing of users could be stimulated by providing incentives and pricing strategies. The assignment of incentives increases the complexity of the system, as incentives should be provided to users who effectively could not perform the action without them. In [43], the solution focussed on the

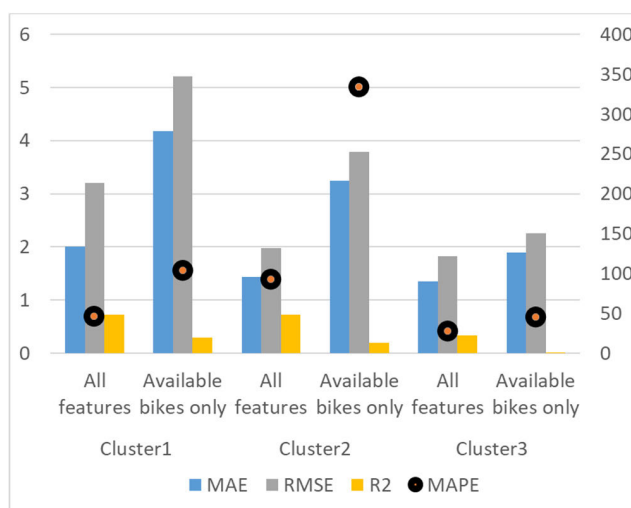


FIGURE 7. Bi-LSTM results comparison using only the number of available bikes as input, compared to using all the features in the dataset. (MAPE values are on the right end).

analysis of the user behaviour, while identifying the typical starting/end points and the travel means, by collecting a large amount of data from a mobile App and processing it in real-time, which is very expensive (this approach permits the computation of the origin-destination matrices, which are also useful for positioning bike racks). An alternative solution would be to provide systematic incentives, for example, to provide a bonus every time the user finds full a bike rack and needs to look for another. This latter condition is avoided with the services providing prediction of bike rack status (e.g., providing bike availability, and thus the number of free slots on the basis of the rack size, as presented in this paper).

From a customer perspective, the prediction of bike rack status is of greater benefit, since it may imply the reduction of their traveling time on a bike ride, and thus a higher quality of service. This aspect has an indirect benefit on the operator's revenues, since a happy customer comes back and exploits the service more often.

## V. CONCLUSION

In this paper, we have proposed a predictive approach and solution for short terms predictions of available bikes in bike-sharing systems with smart stations and with only 3 months of data. Providing reliable predictions can be useful for the development of a relevant number of services both for operators and users.

The solution and its validation have been performed using data collected in bike stations in the cities of Siena and Pisa. The clustering process classified bike racks into 3 clusters, where the representative sensors were identified. The proposed methods use high dimensional time-series data from each representative station and use real-time and forecast weather information as input to perform short-term predictions for the next 15, 30, 45 and 60 minutes. The limited amount of data makes the problem even more difficult for predictive models. State-of-the-art Ensemble Learning and Deep Learning solutions have been compared in order to choose the best one.

The proposed solution demonstrated that when it comes to short-term prediction, even considering the limited amount of data, the Bi-LSTM, Bidirectional Long Short-Term Memory neural network architecture is the most suitable machine learning technique for this problem. The results in terms of Mean Absolute Error in the worst-case have achieved an error of 2 bikes for the 60 minutes prediction on the bike rack.

The most important feature which has been identified by using SHAP analysis is the number of available bikes in all the clusters. The other important features are different from cluster to cluster, but in the top spot there are the temporal information metrics *weekend*, *month*, *day of the week*, meteorological metrics as *temperature* and *pressure* and information about the difference between the number of available bikes at the observation time and past values. By including also this other category of features, results improved if compared to the univariate solution using only the number of available bikes. This analysis has demonstrated

both validity and flexibility of the model which, with the identified features and different relevance obtained good results for different clusters and thus for different behaviours of time series.

The predictions generated by the predictive models have been deployed as an additional feature on Smart City App "Toscana dove, cosa. Km4City" [37] to encourage sustainable mobility.

## ACKNOWLEDGMENT

The authors would like to thank the MIUR, the University of Florence and the different companies involved in co-funding Sii-Mobility national project on smart city mobility and transport. Km4City and Snap4City (<https://www.snap4city.org>) are open technologies and research of DISIT Laboratory (<https://www.disit.org>). Sii-Mobility is based on and has contributed to the Km4City open solution. They also would like to express their thanks to Daniele Cenni for his support.

## REFERENCES

- [1] *World Urbanization Prospects: The 2018 Revision, Online Edition*, United Nations, Dept. Econ. Social Affairs, Population Division, New York, NY, USA, 2018.
- [2] T. Nam and T. A. Pardo, "Conceptualizing smart city with dimensions of technology, people, and institutions," in *Proc. 12th Annu. Int. Digit. Government Res. Conf., Digit. Government Innov. Challenging Times*, 2011, pp. 282–291.
- [3] Y. Zhang and Z. Mi, "Environmental benefits of bike sharing: A big data-based analysis," *Appl. Energy*, vol. 220, pp. 296–301, Jun. 2018.
- [4] A. Li, K. Gao, P. Zhao, X. Qu, and K. W. Axhausen, "High-resolution assessment of environmental benefits of dockless bike-sharing systems based on transaction data," *J. Cleaner Prod.*, vol. 296, May 2021, Art. no. 126423, doi: 10.1016/j.jclepro.2021.126423.
- [5] P. Midgley, "The role of smart bike-sharing systems in urban mobility," *Journeys*, vol. 2, no. 1, pp. 23–31, 2009.
- [6] M. Benchimol, P. Benchimol, B. Chappert, A. De La Taille, F. Laroche, F. Meunier, and L. Robinet, "Balancing the stations of a self-service 'bike hire' system," *RAIRO, Oper. Res.* vol. 45, no. 1, pp. 37–61, 2011.
- [7] D. Chemla, F. Meunier, and R. Wolfler-Calvo, "Balancing a bike-sharing system with multiple vehicles," in *Proc. Congr. Annu. de la Soc. Française de Recherche Opér. et d'aide la Décis. (ROADEF)*, Saint-Étienne, France, 2011, pp. 1–71.
- [8] C. Contardo, C. Morency, and L.-M. Rousseau, *Balancing a Dynamic Public Bike-Sharing System*. vol. 4. Montreal, QC, Canada: Cirrelt, 2012.
- [9] H. R. Gómez Márquez, R. López Bracho, and A. Ramirez-Nafarrate, "A simulation-optimization study of the inventory of a bike-sharing system: The case of Mexico City Ecobici's system," *Case Stud. Transp. Policy*, vol. 9, no. 3, pp. 1059–1072, Sep. 2021.
- [10] J.-R. Lin and T.-H. Yang, "Strategic design of public bicycle sharing systems with service level constraints," *Transp. Res. E, Logistics Transp. Rev.*, vol. 47, no. 2, pp. 284–294, 2011.
- [11] J. E. Froehlich, J. Neumann, and N. Oliver, "Sensing and predicting the pulse of the city through shared bicycling," in *Proc. 21st Int. Joint Conf. Artif. Intell.*, 2009, pp. 1420–1426.
- [12] A. Kaltenbrunner, R. Meza, J. Grivolla, J. Codina, and R. Banchs, "Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system," *Pervasive Mobile Comput.*, vol. 6, no. 4, pp. 455–466, 2010.
- [13] P. Borgnat, C. Robardet, P. Abry, P. Flandrin, J.-B. Rouquier, and N. Tremblay, "A dynamical network view of Lyon's Vélo'v shared bicycle system," in *Dynamics On and Of Complex Networks*, vol. 2. New York, NY, USA: Birkhäuser, 2013, pp. 267–284.
- [14] A. Bargar, A. Gupta, S. Gupta, and D. Ma, "Interactive visual analytics for multi-city bikeshare data analysis," in *Proc. 3rd Int. Workshop Urban Comput.*, vol. 45, New York, NY, USA, 2014, pp. 1–5.
- [15] P. Vogel, T. Greiser, and D. C. Mattfeld, "Understanding bike-sharing systems using data mining: Exploring activity patterns," *Procedia-Social Behav. Sci.*, vol. 20, pp. 514–523, Jan. 2011.

- [16] Y. Li, Y. Zheng, H. Zhang, and L. Chen, "Traffic prediction in a bike-sharing system," in *Proc. 23rd SIGSPATIAL Int. Conf. Adv. Geograph. Inf. Syst.*, Nov. 2015, pp. 1–10.
- [17] L. Chen, D. Zhang, L. Wang, D. Yang, X. Ma, S. Li, Z. Wu, G. Pan, T.-M.-T. Nguyen, and J. Jakubowicz, "Dynamic cluster-based over-demand prediction in bike sharing systems," in *Proc. ACM Int. Joint Conf. Pervas. Ubiquitous Comput.*, Sep. 2016, pp. 841–852, doi: [10.1145/2971648.2971652](https://doi.org/10.1145/2971648.2971652).
- [18] H. Yang, X. Zhang, L. Zhong, S. Li, X. Zhang, and J. Hu, "Short-term demand forecasting for bike sharing system based on machine learning," in *Proc. 5th Int. Conf. Transp. Inf. Saf. (ICTIS)*, Jul. 2019, pp. 1295–1300, doi: [10.1109/ICTIS.2019.8883732](https://doi.org/10.1109/ICTIS.2019.8883732).
- [19] C. Badii, P. Nesi, and I. Paoli, "Predicting available parking slots on critical and regular services by exploiting a range of open data," *IEEE Access*, vol. 6, pp. 44059–44071, 2018. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8430514/>
- [20] C. Badii, E. G. Belay, P. Bellini, D. Cenni, M. Marazzini, M. Mesiti, P. Nesi, G. Pantaleo, M. Paolucci, S. Valtolina, M. Soderi, and I. Zaza, "Snap4City: A scalable IOT/IOE platform for developing smart city applications," in *Proc. Int. Conf. IEEE Smart City Innov.* Nanjing, China: IEEE Press, 2018, pp. 2109–2116. [Online]. Available: <https://ieeexplore.ieee.org/document/8560331/>
- [21] C. Badii, P. Bellini, A. Difino, and P. Nesi, "Smart city IoT platform respecting GDPR privacy and security aspects," *IEEE Access*, vol. 8, pp. 23601–23623, 2020.
- [22] C. Thirumalai and R. Koppuravuri, "Bike sharing prediction using deep neural networks," *Int. J. Informat. Vis.*, vol. 1, no. 3, p. 83, Jun. 2017, doi: [10.30630/ivojv.1.3.30](https://doi.org/10.30630/ivojv.1.3.30).
- [23] X. Wu, C. Lyu, Z. Wang, and Z. Liu, "Station-level hourly bike demand prediction for dynamic repositioning in bike sharing systems," in *Smart Transportation Systems 2019 (Smart Innovation, Systems and Technologies)*, vol. 149. Singapore: Springer, 2019, pp. 19–27.
- [24] N. Boufidis, A. Nikiforiadis, K. Chrysostomou, and G. Aifadopolou, "Development of a station-level demand prediction and visualization tool to support bike-sharing systems' operators," *Transp. Res. Procedia*, vol. 47, pp. 51–58, Jan. 2020, doi: [10.1016/j.trpro.2020.03.072](https://doi.org/10.1016/j.trpro.2020.03.072).
- [25] V. E. Sathishkumar and Y. Cho, "Season wise bike sharing demand analysis using random forest algorithm," *Comput. Intell.*, pp. 1–26, 2020, doi: [10.1111/coin.12287](https://doi.org/10.1111/coin.12287).
- [26] V. E. Sathishkumar, J. Park, and Y. Cho, "Using data mining techniques for bike sharing demand prediction in metropolitan city," *Comput. Commun.*, vol. 153, pp. 353–366, Mar. 2020, doi: [10.1016/j.comcom.2020.02.007](https://doi.org/10.1016/j.comcom.2020.02.007).
- [27] E. A. A. Alaoui and S. C. K. Tekouabou, "Intelligent management of bike sharing in smart cities using machine learning and Internet of Things," *Sustain. Cities Soc.*, vol. 67, Apr. 2021, Art. no. 102702, doi: [10.1016/j.scs.2020.102702](https://doi.org/10.1016/j.scs.2020.102702).
- [28] B. Wang and I. Kim, "Short-term prediction for bike-sharing service using machine learning," *Transp. Res. Procedia*, vol. 34, pp. 171–178, Jan. 2018, doi: [10.1016/j.trpro.2018.11.029](https://doi.org/10.1016/j.trpro.2018.11.029).
- [29] P.-C. Chen, H.-Y. Hsieh, X. K. Sigalingging, Y.-R. Chen, and J.-S. Leu, "Prediction of station level demand in a bike sharing system using recurrent neural networks," in *Proc. IEEE 85th Veh. Technol. Conf. (VTC Spring)*, Jun. 2017, pp. 1–5, doi: [10.1109/VTCSpring.2017.8108575](https://doi.org/10.1109/VTCSpring.2017.8108575).
- [30] S.-H. Choi and M.-K. Han, "The empirical evaluation of models predicting bike sharing demand," in *Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Oct. 2020, pp. 1560–1562, doi: [10.1109/ICTC49870.2020.9289176](https://doi.org/10.1109/ICTC49870.2020.9289176).
- [31] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Statist. Probab.*, 1967, pp. 281–297.
- [32] T. M. Kodinariya and P. R. Makwana, "Review on determining number of cluster in K-means clustering," *Int. J.*, vol. 1, no. 6, pp. 90–95, 2013.
- [33] H. I. Ashqar, M. Elhenawy, and H. A. Rakha, "Modeling bike counts in a bike-sharing system considering the effect of weather conditions," *Case Stud. Transp. Policy*, vol. 7, no. 2, pp. 261–268, Jun. 2019.
- [34] S. Pazdan, M. Kiec, and C. D'Agostino, "Impact of environment on bicycle travel demand—Assessment using bikeshare system data," *Sustain. Cities Soc.*, vol. 67, Apr. 2021, Art. no. 102724.
- [35] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [36] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.
- [37] DISIT Lab. (2021). Toscana Dove, Cosa. Km4City (version 5.5.0) [Mobile App]. Google Play Store. [Online]. Available: <https://play.google.com/store/apps/details?id=org.disit.toscana>
- [38] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Dec. 1997.
- [39] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," *Neural Comput.*, vol. 12, no. 10, pp. 2451–2471, Oct. 2000.
- [40] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, Nov. 1997, doi: [10.1109/78.650093](https://doi.org/10.1109/78.650093).
- [41] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 4768–4777.
- [42] L. Lin, Z. He, and S. Peeta, "Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach," *Transp. Res. C, Emerg. Technol.*, vol. 97, pp. 258–276, Dec. 2018, doi: [10.1016/j.trc.2018.10.011](https://doi.org/10.1016/j.trc.2018.10.011).
- [43] C. Badii et al., "Classification of users' transportation modalities from mobiles in real operating conditions," *Multimedia Tools Appl.*, pp. 1–26, 2021, doi: [10.1007/s11042-021-10993-y](https://doi.org/10.1007/s11042-021-10993-y).



**ENRICO COLLINI** is currently pursuing the Ph.D. degree with the Department of Information Engineering, University of Florence. He is an Engineer. His research interests include deep learning, mobility, and data models.



**PAOLO NESI** (Member, IEEE) is currently a Full Professor with the Department of Information Engineering, University of Florence, and the Chief of the DISIT Laboratory. He has been the chair of several international conferences. He is and has been the coordinator of several research and development multipartner international research and development projects. His research interests include machine learning, massive parallel and distributed systems, physical models, the IoT, mobility, big data analytic, semantic computing, formal model, and machine learning.



**GIANNI PANTALEO** is currently an Aggregated Professor with the Department of Information Engineering, University of Florence. He has been a coordinator of a number of WP in international research and development projects. His research interests include knowledge engineering, the IoT, visual analytics, mobility, NLP, and semantic computing.