Sustainable Cities and Society 67 (2021) 102702

Contents lists available at ScienceDirect

Sustainable Cities and Society

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

Intelligent management of bike sharing in smart cities using machine learning and Internet of Things

El Arbi Abdellaoui Alaoui [a,b,*], Stephane Cedric Koumetio Tekouabou [c,d]

[a] EIGSI-Casablanca, 282 Route of the Oasis, Casablanca, Morocco
[b] My Ismail University, Morocco
[c] Center of Urban Systems (CUS), Mohammed VI Polytechnic University (UM6P), Hay Moulay Rachid, 43150 Ben Guerir, Morocco
[d] Laboratory LAROSERI, Department of Computer Science, Faculty of Sciences, B.P. 20, 24000 El Jadida, Morocco

ARTICLE INFO

Keywords:
Bike-sharing system (BSS)
Management
Prediction
Smart cities
Internet of Things (IoT)
Regression
Ensemble models

ABSTRACT

Global ecological requirements are pushing city actors to opt for ecological solutions at all levels, including urban mobility. More sustainable Bike-sharing systems (BSS) have become an indispensable part of the transport offer by world's major metropolis. Like any computerized service system, they generate voluminous and complex data that the use of which is essentially limited to the management and operation of the system. The movements made by system users can provide valuable information on many aspects of urban life including the spatial and temporal dynamics of travel in the city, on the place of the bicycle among other modes of transport, or on the distribution of territorial and social inequalities in geographical space. In this paper, we study the problem of intelligent management of shared bicycle systems. Indeed, the management of these systems faces many optimization problems in its procession. Thus, to improve the BSS user's satisfaction, it's useful to inform the actors/users in this system about the state of bike sharing for a station. For this, we propose an approach that integrates in these systems both the new IoT for smart city technologies and machine learning in order to facilitate the task of management, availability and profitability. In addition, we propose an automatic management system capable of predicting the number of bikes shared per hour, day or month by taking several dynamic parameters. Simulation results carried out on real data from London's bike sharing system demonstrate the effectiveness of the proposed model.

1. Introduction

Bike-sharing systems have become a key element in urban transport policies over the last decade, as evidenced by the recent explosion in the number of bicycles in circulation in the world's major cities. By providing affordable access to bicycles, they should actively participate in the development of alternatives to motorized vehicles for urban travel and contribute to the reduction of air pollution (Cao & Shen, 2019; Leister, Vairo, Sims, & Bopp, 2018), noise levels and congestion problems affecting the world's major cities. One of the keys to the success of the bike sharing system is its ease of use. For the majority of systems, a user has the option of picking up or dropping off a bike at one of the many stations throughout the city in a fully automated manner. They are known as "docked BSS". In recent years, an alternative based on dockless BSS has emerged in several metropolises, although remaining less widely used than the first one. Instead of a fixed station as in the case of docked BSS, dockless BSS are centered on bike with a lock that is usually integrated into the frame without fixed station allowing the user the flexibility to pick up and drop off the bike wherever he wants (Chen, van Lierop, & Ettema, 2020; Ji, Ma, He, Jin, & Yuan, 2020). In either case, the user can pay a daily, weekly, monthly or annual subscription available immediately at the station or mobile application. Payment can be made by direct debit or by card. Each journey entitles the user to a period of free or very low cost use. Generally this period varies between 30 minutes and 1 hour, beyond which the price increases significantly, encouraging short trips. However, the number of journeys is unlimited during the subscription period, allowing users to make several journeys in a row, provided that they can leave the bicycle at a station. Thus, actual metropolis bike sharing program (BSP) consists both of docked and dockless BSS. Predicting the number of shared biked over a time in such a metropolis is a challenge for the BSP (Bikeshare, 2013; Garcia-Gutierrez, Romero-Torres, & Gaytan-Iniestra, 2014).

* Corresponding author.
E-mail address: abdellaoui.e@gmail.com (E.A. Abdellaoui Alaoui).

https://doi.org/10.1016/j.scs.2020.102702
Received 12 April 2020; Received in revised form 14 November 2020; Accepted 29 December 2020
Available online 9 January 2021

BSS are a new generation of traditional bike rentals where the whole process of joining, renting and returning bikes has become automatic. Thanks to these systems, the user can easily rent a bike from a particular position and return to another position. Currently, there are about 500 bike-sharing programs worldwide, with over 500,000 bikes shared. Today, there is great interest in these systems because of their important role in traffic, environmental and health issues (Cao & Shen, 2019).

In addition to the interesting real-world applications of bike-sharing, the characteristics of the data generated by these systems make them attractive for research in sustainable urban planning (Lu, An, Hsu, & Zhu, 2019). In contrast to other transport services such as bus or metro, travel time, departure and arrival positions are explicitly recorded in these systems. This features transform the BSP into a virtual sensoring network that can be used to detect mobility in the city using new technologies such as the Internet of Things (IoT). Indeed, BSP and the IoT are strongly linked. Currently, most BSP use 2Gs for networking, but these systems will gradually move to narrow-band networks of the Internet of Things (NB-IoT) in the future (Chen, Miao, Hao, & Hwang, 2017). Several bike-sharing companies have started using the NB-IoT (Sinha, Wei, & Hwang, 2017). Recently a study has been conducted by Zguira, Rivano, and Meddeb (2018) to facilitate the collection of BSP data through the IoT. They proposed an efficient routing protocol, IoB-DTN, based on data aggregation that applies the Delay Tolerant Network (DTN) for IoT applications which collect data from sensor networks based on urban bike sharing systems (Zguira et al., 2018). Therefore, most significant events in the city should be detected by monitoring this data using the services of a reliable network.

BSP typically aim to reduce traffic congestion, noise and air pollution (Bullock, Brereton, & Bailey, 2017; Cao & Shen, 2019) by providing free/affordable access to bicycles for short-distance travel in an urban area as opposed to motorized vehicles. The number of users on a given day can vary considerably for these systems. The ability to forecast the number of hourly users can enable entities such as companies and governments that oversee these systems to manage them more efficiently and cost-effectively. Predicting shared bikes is also very important for urban planning and business model for companies investing in shared bike programs in urban areas (Bikeshare, 2013). Our goal is to use and optimize machine learning based models that effectively predict the total number of shared bikes that will be used during a given hour, a given month, a given season, etc., using the information available on that hour/day within the bike system, taking into account several factors (Eren & Uz, 2019; Mattson & Godavarthy, 2017). Thus, it is useful to inform the actors/users in this system about the bike sharing status for a station or around a user's location.

In this paper, we propose a combination of IoT and machine learning techniques through regression ensemble methods to optimize the management of self-service shared bike systems in smarts cities that we also want to be green and sustainable cities. This type of combination has already shown good performance in the case of intelligent parking garages (Tekouabou, Cherif, & Silkan, 2020). Indeed if the IoT allows to collect a large variety of data from the different dynamic sources associated to our BSS, after data cleaning and transformation, the robust ensemble methods could allow to predict the number of shared bikes per hour, per day, per month, etc., thus allowing to optimize and automate the efficient management of the system. We now have a volume of data that allows us to study the external environment with precision. The city is fully digitized and the amount of data produced in the cities will continue to double every 2 years. These data are all integrated and processed automatically by the prediction algorithms. They enable predictions to be made such as the fill rate of a bike share station, system or program; the time it takes to find a parking space on the road, public transit trips and even the comfort or stress felt by pedestrians.

The key contributions of this work are:

- The integration of an IoT-based system in smart cities and particularly in the case of bike sharing;
- The global system uses ensemble methods to predict the number of bike sharing in smart city;
- The optimization of different ensemble-based models for predicting number of bike sharing in smart city;

The rest of this paper is organized as follows: In Sections 1 and 3.1 we first present respectively previous work on bike sharing and bake sharing approach for smart cities. We then describe the proposed method to predict the number of bike shared in Section 4 before analyzing the results obtained in Section 5. Finally, we discuss the results of our model and its limitations in Section 6 and conclude our study in Section 7.

2. Related work

In parallel with the explosion of bike sharing programs around the world, research on bike sharing has seen a burgeoning interest from transportation researchers and beyond. The synthesis of the literature on bike sharing systems by Fishman et al., first in 2013 (Fishman, Washington, & Haworth, 2013) and again in 2015 (Fishman, 2016), reveals both the wealth of scientific publications on the subject and the diversity of research topics covered. These studies thus examine the many questions raised by this new mode of transport, whether they are:

(i) technical: does the system work properly? is it optimal? can it be improved and how?,

(ii) Social: who uses BSS and why? How to improve access to BSS?

(iii) in relation to transport: what is the impact of BSS systems on other modes of transport?

(iv) or dealing with economic aspects, such as what the activity of bike sharing program tells us about the economic activity of a city.

This enthusiasm can be explained by the need, both for political decision-makers and the companies managing such systems, to know and describe BSS systems in order to develop and support the use of bike sharing in the city.

In addition, unlike traditional transportation systems that require the establishment of often cumbersome and costly infrastructure (like using sensors or field surveys in order to obtain partial system data), the technologies employed in the third-generation BSS currently in operation provide rapid access to digital data on multiple aspects of bike sharing. This data, provided by the operators of "JCDecaux developer" platform Jcdecaux developer (2020) allows, for example, access to the status of the stations of all the systems managed by cyclocity in real time. It has fostered collaboration between specialists in socio-economic issues, who until then had been dealing with these problems, and researchers in data analysis (signal processing, data mining, etc.), in order to find answers to the problems mentioned above, in data sets that can be large and require sophisticated techniques.

2.1. Bike sharing as a mode of transport

Beyond surveys carried out by institutional bodies (Transportation, 2013) or the operators themselves (Bikeshare, 2013; London, 2013) consisting of qualitative or quantitative descriptions of the use of the system and its management, the place of bike sharing among urban travel modes has been the subject of many studies, seeking to understand the reasons for the success of BSP. They relate to the analysis of a particular system, such as Montreal (Morency, Trepanier, & Godefroy, 2011), Hangzhou (Shaheen, Zhang, Martin, & Guzman, 2011), Bangkok (Ueasangkomsate, 2020), London (Noland & Ishaque, 2006) or Dublin (Tuama, 2015), or comparative studies between several systems in North America, Western Europe and Asia (Lohry, 2020; Midgley, 2009; Parkes, Marsden, Shaheen, & Cohen, 2013; Shaheen, Guzman, & Zhang, 2010).

The cohabitation of bike shared with other modes of transportation and particularly the effects of BSP on the reduction of car use have been scrutinised: urban agglomerations are seeking to promote cycling as part of policies to control energy consumption and greenhouse gas emissions, mainly from motorised transport (DeMaio, 2009). Several studies have shown the low impact of bike sharing on vehicle use (Bikeshare, 2013; Midgley, 2011; Murphy, 2010; Yang, Haixiao, & Qing, 2020), through field surveys. The vast majority of bike shared user's are those of soft transport modes (public transport, walking, etc.). The role of BSP in the promotion of cycling in cities has also been studied through the transition between shared and personal bikes (Castillo-Manzano, Castro-Nuno, & Lopez-Valpuesta, 2015) or factors influencing the practice of shared bikes (Kim, Shin, Im, & Park, 2020). The health benefits of cycling induced by the use of BSS have also been the subject of several studies (Doorley, Pakrashi, & Ghosh, 2015; Fishman, Washington, & Haworth, 2012; Rojas-Rueda, de Nazelle, Tainio, & Nieuwenhuijsen, 2011; Woodcock, Tainio, Cheshire, O'Brien, & Goodman, 2014), and the London (Woodcock et al., 2014) example illustrates the overall positive impact of cycling in cities on certain categories of population. The characterization of users, and their motivations, are also at the origin of many studies, mainly conducted by sociological researchers. This work has been carried out either through individual surveys, as in Montreal (Bachand-Marleau, Lee, & El-Geneidy, 2012; Fuller et al., 2011), or data analysis methods, like in London (Ogilvie & Goodman, 2012). Topics include understanding the behaviors involved in the relationship between users and bike-sharing (Bordagaray, Ibeas, & dell'Olio, 2012), studying the social characteristics of BSS users such as gender (Beecham & Wood, 2014) or social level (Hoe & Kaloustian, 2014), and the factors that differentiate between regular cyclists with their own bikes and BSS users (Buck et al., 2013).

2.2. Bike sharing data analysis

The analysis of BSP through data is also the source of many studies, often multi-disciplinary. Bike sharing is characterized by temporal and spatial analyses, for example through the classification of bicycle stations, as in Paris (Etienne & Latifa, 2014; Nair & Miller-Hooks, 2011; Randriamanamihaga, Côme, Oukhellou, & Govaert, 2014), Barcelona (Froehlich, Neumann, & Oliver, 2008, 2020; Kaltenbrunner, Meza, Grivolla, Codina, & Banchs, 2010) or London (Lathia, Ahmed, & Capra, 2012) to name the main ones. Comparative multi-city approaches have also been implemented (Austwick, O'Brien, Strano, & Viana, 2013; Bouveyron, Côme, & Jacques, 2015; O'brien, Cheshire, & Batty, 2014; Sarkar, Lathia, & Mascolo, 2015), and reflect the strong similarity between all the systems, particularly with regard to the spatial and temporal distribution of activities. Most of these studies highlight relationships between the use of bicycles, the time of day and a geographical and socio-economic description of the city.

Several efforts have also been made, particularly by research groups from the exact sciences, to forecast demand in order to plan the system and anticipate its development (where to locate stations in the city, how many locations per station, etc.), both through theoretical (Bortolussi & Hillston, 2015; Chen & Sun, 2015; Dell'Olio, Ibeas, & Moura, 2020; García-Palomares, Gutiérrez, & Latorre, 2012; Krykewycz, Puchalsky, Rocks, Bonnette, & Jaskiewicz, 2010; Lin & Yang, 2011; Saharidis, Fragkogios, & Zygouri, 2020) and data-driven approaches as in Lyon (Borgnat et al., 2011, 2013), in Hangzhou (Woodcock et al., 2014) or in Montreal (Faghih-Imani, Eluru, El-Geneidy, Rabbat, & Haq, 2014). The latter problem was recently the subject of a competition held in conjunction with a conference on data mining, reflecting the popularity of this topic (Subramaniyan, 2020). Other issues related to the smooth day-to-day operation of the system (bike maintenance, station balancing, etc.) were also studied, once again through the development of theoretical models (Nair & Miller-Hooks, 2011; Sayarshad, Tavassoli, & Zhao, 2012; Schuijbroek, Hampshire, & Van Hoeve, 2017; Vogel, 2020; Vogel, Saavedra, & Mattfeld, 2020; Waserhole, Jost, & Brauner, 2013) and the analysis of real data (Nair & Miller-Hooks, 2011; Schuijbroek et al., 2017; Vogel, Greiser, & Mattfeld, 2011).

Table 1 provides an overview of publications dealing with data analysis for the study of bike sharing, classified by city and research theme, for a selection of cities.

2.3. System prediction

In the literature, we find some work based on machine learning mechanisms to determine the resources in the BSS. In Table 2 we have grouped together works dealing with resource prediction for Bike sharing.

3. Bike sharing in smart city

3.1. Concept of smart city

Smart cities are a concept developed in recent years that aims to offer solutions to improve the management of cities, particularly through the use of new technologies (Bouzguenda, Alalouch, & Fava, 2019; Yigitcanlar et al., 2019). The urban space is seen as a complex superposition

Table 1
Summary of the related work of dealing with data analysis for the study of bike-sharing systems, classified by city and research topics, on a selection of the most studied cities.

City	Spatial and temporal analysis	Regulation	Flow prediction	Classification of stations	Others
Barcelona	Kaltenbrunner et al. (2010), O'brien et al. (2014)			Froehlich et al. (2008), Froehlich et al. (2020), Sarkar et al. (2015)	
Dublin					Yoon, Pinelli, and Calabrese (2012)
Hangzhou			Xu, Ying, Wu, and Lin (2013)		
London	O'brien et al. (2014)			Bouveyron et al. (2015), Lathia et al. (2012), Sarkar et al. (2015), Tribastone, Clark, Gast, Gilmore, and Reijsbergen (2014), Austwick et al. (2013)	Ogilvie and Goodman (2012)
Lyon	Borgnat et al. (2011), O'brien et al. (2014)		Borgnat et al. (2011), Borgnat, Fleury, Robardet, and Scherrer (2020)	Borgnat et al. (2013), Bouveyron et al. (2015)	Jensen, Rouquier, Ovtracht, and Robardet (2010)
Montreal	O'brien et al. (2014)		Faghih-Imani et al. (2014)		
Paris	Etienne and Latifa (2014), O'brien et al. (2014)	Nair, Miller-Hooks, Hampshire, and Bušić (2013)		Bouveyron et al. (2015), Etienne and Latifa (2014), Randriamanamihaga et al. (2014)	
Vienna	O'brien et al. (2014)	Vogel et al. (2011)		Vogel et al. (2011)	Test
Washington	O'brien et al. (2014)	Schuijbroek et al. (2017)		Austwick et al. (2013)	Fanaee-T and Gama (2014)

Table 2
Summary of related work on usage prediction.

Ref	City	Timespan	Algorithm	Performance measures	Best performances
Fanaee-T and Gama (2014)	Washington DC	2 years	Weka's regressors	-Relative Absolute Error -Root Relative Squared Error	29.98% 39.27%
Yoon et al. (2012)	Dublin	27 days	Modified ARIMA	Root Mean Square Error	5 min: 0.91 60 min: 3.47
Froehlich et al. (2020)	Barcelona	13 weeks	Bayesian network	Relative Absolute Error	0.08
Borgnat et al. (2011)	Lyon	2 years + 8 months	Linear Regression	Mean Relative Error:	12%
Kaltenbrunner et al. (2010)	Barcelona	7 weeks	Auto-Regressive Moving Average	Mean Absolute Error	1.39

of infrastructures grouping together municipal services such as transport, water, energy and energy supply, or systems communications (Sharifi, 2020). Above these infrastructures, a new information layer is put in place, making it possible to obtain, with the help of sensors, real-time information on many aspects of the systems in order to optimise maintenance, inform citizens and, in a second stage, understand the functioning of these systems in order to improve them. As Jeremy Rifkin explains (Rifkin, 2011), smart grids, or smart energy grids, propose a new paradigm in energy management, shifting from an approach in which a company produces and delivers energy to consumers, to another model in which each household produces energy and distributes it to the grid. The concept of Smart Cities extends this paradigm to the city level, offering inhabitants the means to become responsible players in the management of the city. Due to the large volume of data generated, their variety and speed of acquisition, issues related to smart cities are part of the datamass (Sharifi, 2020).

The first smart city projects still remain evasive about their aims, contenting themselves with general ideas. Fig. 1 shows an example of a smart city. The vision of the smart city is broken down into four ideas (Sharifi, 2020):

- The taking into account environmental issues and clean energy constraints;
- The networking of the actors among themselves;
- The transition from property to use;
- The integration of new technologies (like the IoT).

Research on smart cities requires a very broad set of skills given the diversity of problems encountered in cities. It also requires the development of techniques for extracting knowledge adapted to urban data.

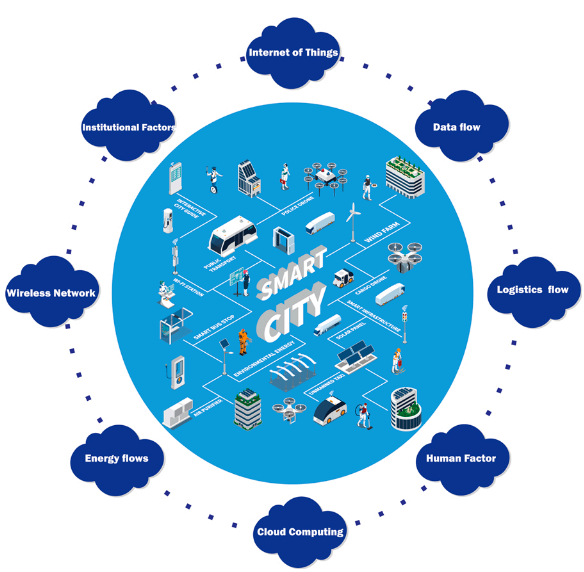


Fig. 1. Fundamental concept of the smart city.

These data are very often found in the form of networks, energy or communication networks. Fig. 1 shows, for example, the main intervened acts in the design of a smart city. The sensors also form networks and can exchange information according to their distance in order to better control, for example, urban pollution (Sharifi, 2020). This particular structure, found in many systems, is the subject of a new theory that has been developing in recent years: the science of complex networks.

3.2. Bike sharing program (BSP)

A BSP or bike sharing, or self-service bike sharing consists of making bikes available to the public, free of charge or not. This type of mobility service allows people to travel locally, mainly in urban areas, while preserving the environment. This bike rental is a form of collaborative consumption and thus removes three obstacles to cycling: parking at home, theft and maintenance of one's personal bike. While they are equally good with and without docks, they have the particularity of being centered on a bike.

3.2.1. Bike

The concept of bike sharing is a neologism of shared bikes and corresponds to a service that allows an individual to move around easily without the need to have a personal bike. The bike is an energy-efficient, safe, CO_2-neutral and space-saving transportation mode (Cao & Shen, 2019). It has a small ecological footprint (if used). In urban areas, for short trips, it is a good alternative to the car. For longer trips or to get to work, always in an urban environment, it is an excellent complement to public transport, as it multiplies the area served. Although at the beginning of the 21st century, the majority of the BSS were docked, nowadays, the BSP (see Fig. 2) are made up of both docked (Fig. 2a) and dockless (Fig. 2b) BSS which have emerged recently in several cities such as London, New York, San Francisco, Beijing, etc.

3.2.2. Docked BSS

A docked bike-sharing system consists of bikes, roads, and fixed stations, as shown in Fig. 2a. Unlike dockless BSS, the station plays a very important role in this type of BSS, which is still very successful today and particularly in London city.

3.2.3. Station in docked BSS

The main characteristic of docked BSS is based on fixed stations where users can pick up or drop off the shared bike on a self-service basis. In such system, each station is composed of a transactional terminal and several bike anchorage points. Each transactional terminal is equipped with a radio identification reader (RFID) and sensors that transmit information to the transactional terminal. However, since this obligation to find a station is a constraint for users and also affects the flexibility of the system, it has led to the emergence of dockless BSS as in the case of the London BSP.

3.2.4. Dockless BSS

According to Sherriff, Adams, Blazejewski, Davies, and Kamerāde

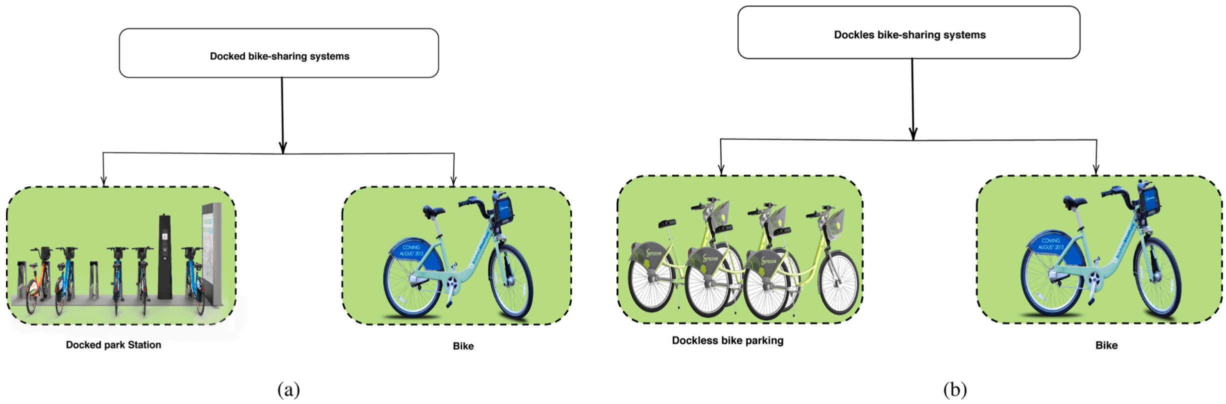


Fig. 2. Example of Bike Sharing Program (BSP) consisting of: (a) Docked BSS and (b) Dokless BSS.

(2020), the utility of a bike share service is dependent on accessibility and availability of the bikes. The location of bikes is therefore an important consideration. Like docked BSS, a dockless BSS also makes bikes available to the public, free of charge or not. It allows people to travel locally, mainly in urban areas. Dockeless BSS consists of bikes and roads without fixed parking stations as shown in Fig. 2b. The latter are also known as floating bike or fourth generation BSS. Thus, dockless bike rental systems consist of a bike with a lock that is usually integrated into the frame and does not require a docking station (Ji et al., 2020; Lazarus, Pourquier, Feng, Hammel, & Shaheen, 2020).

3.3. Factors affecting bike sharing in smart sities

Many factors could affect bike sharing in smart cities and can be: collective (economic, cultural and social), individual (physical conditions, personal constraints, etc.) or meteorological factors. They play an important role in the shared bike and should be considered when promoting paratransit cycling as a mode of travel. Several papers have been published with great interest for the time of day, day of the week, month, etc. (Zhang, Wen, Qiu, Wang, & Abbas, 2019). However, as weather issues are frequently cited as a major impediment to use bike-sharing. including weather condition in BSP designing is one of the issues of this paper.

4. Methodology

In this section we propose an approach to predict the total number of bike shares. Fig. 3 describes an overview of the proposed methodology. This approach is mainly based on the introduction of different steps of the supervised regression process. It consists first of all of a new enrichment method to improve the quality of the data and to take into account other relevant factors (e.g. environmental factors). Secondly, there is the implementation of machine learning algorithms to predict the number of shared bikes. Thirdly, we adjust the parameters of each model. Finally, we propose adequate evaluation tools based on regression performance metrics.

4.1. Smart system for Bike-sharing management

An intelligent bike sharing system can be defined as a machine that integrates a server (computer) connected to the Internet that can collect and analyse data and communicate with other systems. It is therefore a classic intelligent system adapted to the context of bike sharing (see Section 3.2.2). Such a system must therefore effectively perform three functions: data collection, analysis and transmission. The intelligence embedded in these systems is part of the development of the IoT where almost everything can be provided in the form of unique identifiers, and allows data to be transferred automatically over a network without human-human or human-machine interaction. Nowadays, one of the

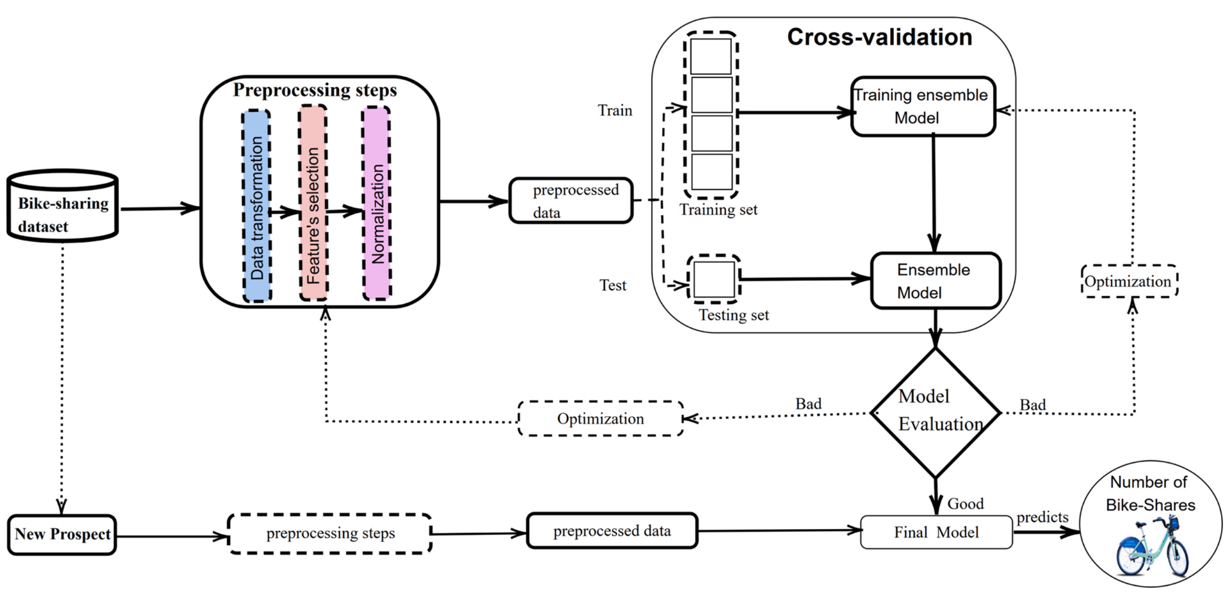


Fig. 3. Global ensemble-based system for real-time average number of bike-share prediction.

most important factors of intelligent systems is the inclusion of advanced software systems based on artificial intelligence that make it possible to anticipate a user's needs in order to better satisfy them. This artificial intelligence is based on the use of machine learning algorithms. In this part, we propose a process for building a machine learning model based on ensemble methods to optimize the prediction of bike sharing in an smart city. The proposed approach consists of three steps illustrated in Fig. 4.

4.2. Data collection and prepocessing

The first step to achieve best prediction according to flowchart 3 is therefore the collecting and the aggregation of sufficient data, for solving the problem. The second step is the pre-processing of the collected data, i.e. a reduction to just what we need, as well as their cleaning, transformation and standardization. The goal of this step is to increase the accuracy of the model and to optimize the execution time and learning as well as its size. More details about the dataset is in Section 5.1.

4.3. Ensemble-based modeling

The final goal of a machine learning work is to build a prediction model. Indeed, one of the main challenges of machine learning is to design efficient regression systems based on a set of representative examples of a data population. Among the different approaches to address this type of problem, combining a set of weak individual regressors to form a single regression system called the set of regressors has attracted growing interest from the scientific community. Recent work of research has shown that certain principles of combining classifiers are particularly effective, such as Boosting, Bagging, Random Subspaces, or more recently Random Forests. The efficiency of regressor combinations is mainly based on their ability to take advantage of the complementarities of the individual regressors, with the aim of improving performance as much as possible by generalizing the set (Tay, Chui, Ong, & Ng, 2013).

The learning machine offers many different models to build a relationship between two magnitudes, the X input and the Y output, but the ideal is to build the most synthetic relationship possible. The simpler it is, the easier it is to interpret. Ensemble methods are good at finding relationships between specific values of X and Y. Indeed, ensemble (or aggregation) methods for statistical learning algorithms are based on the idea of combining the predictions of several predictors (or classifiers) for better generalization and to compensate for the possible shortcomings of individual predictors, such as support vector machine regressor (SVMR) (Mavroforakis & Theodoridis, 2006) and linear regression (LR) (Coppi, D'Urso, Giordani, & Santoro, 2006). In general, there are two families of such methods:

- Bagging methods (Chen & Ren, 2009) where the principle is to average several predictions hoping for a better result following the reduction of variance of the mean estimator. We find in this family the following algorithms: Random Forest Regressor (RFR), Bagging Regressor (BR).
- Boosting methods (Freund, Schapire, & Abe, 1999) where the parameters are iterative adapted to produce a better mix. We find in this family the following algorithms: XGBoost Regressor (XGBR), AdaBoosting Regressor (ABR).

In the following we will explore each of these algorithm classes.

4.3.1. Random forest regressor

Random Forest are a combination of decision trees, where each tree depends on the values of an independently sampled random vector with the same distribution for all trees in the forest. The generalization error of a forest of trees depends on the strength of the individual trees in the forest and the correlation between them. Using a random selection of characteristics to divide each node gives error rates that compare favorably to other ensemble methods (Géron, 2019). The algorithm 1 describes the Random Forest principle (Breiman & Cutler, 2005; Cutler, Cutler, & Stevens, 2012).

Algorithm 1. Random Forest Regressor (RFR) Breiman and Cutler (2005), Cutler, Cutler, & Stevens(2012)

Input • x the observation to predict;
• d_n the observation;
• B the number of Trees;
• $m \in N$ the number of candidate variables to cut a node.

Output $h(x) = \frac{1}{B}\sum_{k=1}^{B} h(x, y_k)$

1: **for** $k = 1$ to B **do**
2: Draw a bootstrap sample in d_n
3: Construct a Classification And Regression Tree (CART) Razi and Athappilly (2005) on this bootstrap sample, each cutoff is selected by minimizing the cost function of CART over a set of m randomly selected variables among the p. We note $h(., k)$ the built tree.
4: **end for**

where $h(x, y_k)$ is a collection of predictors per tree where y_k is a string of i.i.d. random variables. The random forest predictor is obtained by aggregating this collection of trees.

4.3.2. Bagging regressor (BR)

Bagging (Breiman, 1996; Chen & Ren, 2009), meaning "bootstrap aggregating", was one of the first methods proposed for the creation of a set of regressors. The Bagging technique redefines the learning set for each regressor. It consists in generating several versions of the same training database (called bootstraps) and using each of them to train a basic estimator using the same training algorithm. The different bootstraps are built by random draws with delivery of examples of the original training data set. Thus, the outputs of the classifiers are then combined by majority vote or any other merger rule Breiman (1996).

Algorithm 2. Bagging Regressor (BR) Breiman (1996)

Input • x the observation to predict;
• d_n the observation;
• B the number of Trees;
• $m \in N$ the number of candidate variables to cut a node.

Output the estimator $\widehat{m}(x) = \frac{1}{B}\sum_{k=1}^{B} \widehat{m}_k(x)$

1: **for** $k = 1$ to B **do**
2: Draw a bootstrap sample d_n^k in d_n
3: Adjust the regressor on this bootstrap sample $\widehat{m}(x)$.
4: **end for**

4.3.3. XGBoost regressor (XGBR)

In this section, we will present the XGBoost algorithm proposed by Chen and Guestrin. The name XGBoost is an abbreviation of "eXtreme Gradient Boosting". This algorithm is a continuation of GBM (bradient boosting machines). The interest of XGBoost lies in the correction of a major problem of Gradient Boosting Tree, namely the fact of considering the potential loss for each possible separation (split) at a given node to create a new branch (Chen, 2020). Indeed, in the case where we have a lot of explanatory variables, there are a large number of possible splits, which leads to high computation times. Fig. 5 shows the algorithm process.

As depicted in Fig. 5, at each iteration of the gradient boosting algorithm, the residual will be used to correct the previous predictor so that the specified loss function can be optimized. As an improvement, the regularization is added to the loss function to establish the objective function in XGBoost measuring the performance of the model Θ Zhang et al. (2018). Such performance Θ is given by Eq. (1) where: τ denotes the trained parameters from the given data; Φ is the training loss function, such as square loss or logistic loss, which measures how well the model fits on training data and Ω is the regularization term Zhang et al. (2018).

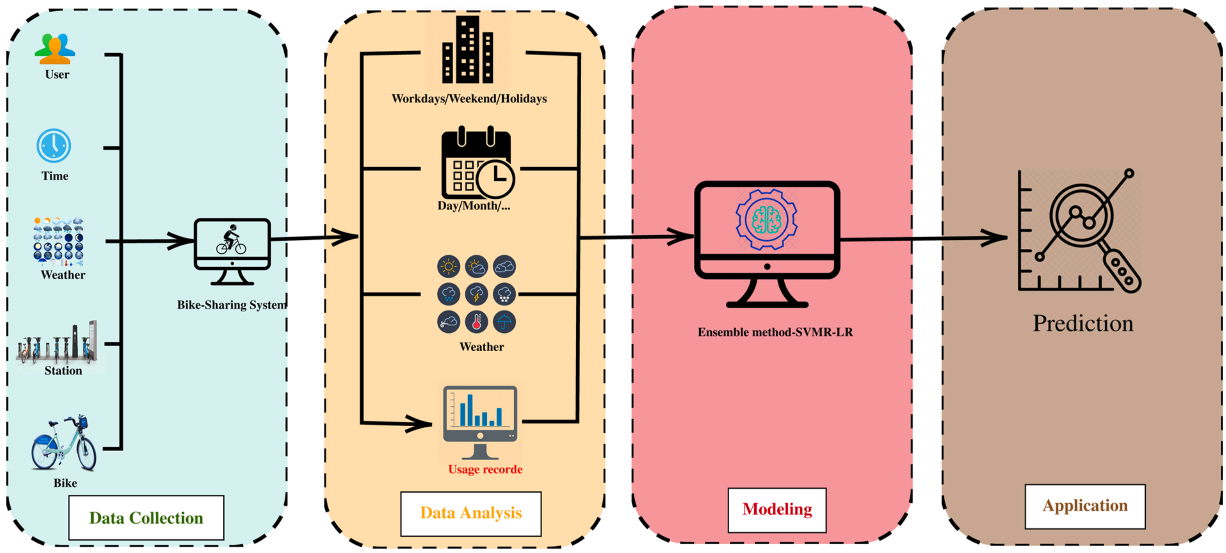


Fig. 4. Smart system for bike-sharing management.

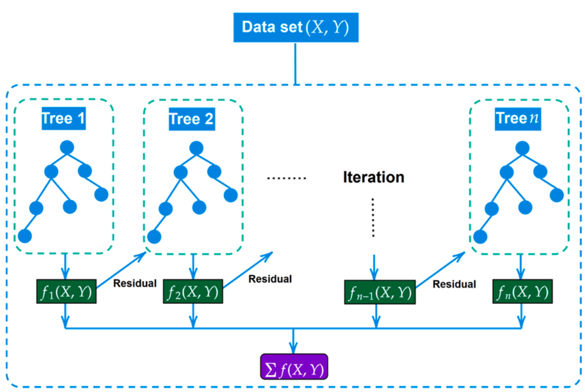


Fig. 5. Algorithm process of XGBoost.

$$\Theta(\tau) = \Phi(\tau) + \Omega(\tau) \tag{1}$$

The algorithm can independently determine the types of loss functions used for model evaluation. To reduce the risk of overfitting, an additional adjustment term is added to the model. Knowing that the base models f_m consist of decision trees, we can write the output of the ensemble model $\hat{y}_i$ from a space W, which is voted or averaged over m base trees by Eq. (2).

$$\hat{y}_i = \sum_{i=1}^{m} f_m(x_i), \quad f_m \in W, \tag{2}$$

For n predictions, the objective function at the t^{th} iteration can be written by Eq. (3) below.

$$\Theta(t) = \sum_{i=1}^{n} \Phi(y_i, \hat{y}_i) + \sum_{k=1}^{t} \Omega(f_k) \tag{3}$$

Given objective function of Eq. (3), the corresponding output $\hat{y}_i^{(t)}$ of the ensemble model will be given by Eq. (4) as follows:

$$\hat{y}_i^{(t)} = \sum_{i=1}^{m} f_m(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i) \tag{4}$$

The regularization term $\Omega(f_k)$ for a decision tree is defined by Chen (2020) as follows:

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \sum_{i=1}^{m} \omega_j^2 \tag{5}$$

From Eq. (5), γ denotes the complexity of each leaf, T denotes the number of leaves in a tree, λ denotes a parameter to scale the penalty, and ω is the vector of scores on the leaves. Then, the first-order along with the second-order Taylor expansions are taken to the loss function in XGBoost. More details about the process of XGBoost algorithm are given by Chen (2020), Zhang et al. (2018).

4.3.4. AdaBoosting regressor (ABR)

Adaboost was able to solve several optimization problems that were causing problems with the boosting algorithms. It takes up the general principle of boosting by constructing a linear combination of hypotheses, designed by iterative executing a weak leaner algorithm, on probability distributions calculated on the basis of the results of the previous iterations (Freund & Schapire, 1997, Freund et al., 2020). This is to allow the learner to focus his or her attention on examples that are difficult to learn (Mishra, Mishra, & Santra, 2017). Algorithm 3 describes the principle of the Adaboost algorithm.

Algorithm 3. AdaBoosting Regressor (ABR) Freund and Schapire (1997), Freund et al. (2020)

$$g_0(.) = argmin_c f(x) = \frac{1}{n}\sum_{m=1}^{M} L(y_i, c)$$

Input •x the observation to predict;
•$d_n = (x_1, y_1), \ldots, (x_n, y_n)$ the sample;
• h the weak rule;
•λ a regularization parameter such as $0 < \lambda \leq 1$.
•M the number of iterations.;
Output the estimator $\hat{g}_M(x)$
1: *Initialization*:

$$g_0(.) = argmin_c f(x) = \frac{1}{n}\sum_{m=1}^{M} L(y_i, c)$$

2: **for** $m = 1$ to M **do**
3: Calculate the opposite of the gradient $\frac{\partial}{\partial y} L(y, g)$ and evaluate it at points $g_{m_1}(x_i)$:

(*continued on next page*)

(*continued*)

$$U_i = -\left[\frac{\partial l(y_i, g_m(x_i))}{\partial g(x_i)}\right]_{g(x_i)=g_{m-1}(x_i)}, i = 1, \ldots, n.$$

Adjust the weak rule on the sample $(x_1, U_1), \ldots, (x_n, U_n)$, we notice h_m the rule thus defined.
Update: $g_m(x) = g_{m-1}(x) + \lambda h_m(x)$
end for

5. Experiments and results analysis

5.1. Bike sharing dataset

The used dataset in our experiments comes from the repository of the city of London in England. The data come from three different sources: the TfL government website[1] that shares London's bike sharing data, the weather site[2] on which the meteo data for the same period were collected and finally the government site[3] allowing to obtain working days, holidays and weekends. Data were collected from 04/01/2015 to 03/01/2017 for these three sources.

These information are publicly available on the links associated with each of the data sources. The data from the three sources was then aggregated according to time, day and weather information and analyzed in the 5.4.1 section. The type of the attributes as mentioned in the third column of Table 3, corresponds to the type of original data collected for each of these attributes. Thus, prior to this analysis, some features have been transformed as follows[4] :

- Season: there are 4 seasons corresponding to: spring coded by (0), summer coded by (1), fall coded by (2) and winter coded by (3).
- Working day: if the day is neither the weekend nor a public holiday, the value is 1. Otherwise, 0;
- Weather situation: we have eight cases; Clear or Few clouds is coded by (1), Partly cloudy is coded by (2), Fragmented clouds is coded by (3), Fog more cloudy is coded by (4), Rain is coded by (7), Heavy rain or ice pallets plus thunderstorm is coded by (10), Snowfall is coded by (26) and Icing fog is coded by (94).

After codification, all features V_{ij} were normalized according to the formula of Eq. (6) to minimize the impact of variable scales on predictor performance.

$$V_{ij} \longleftarrow \frac{V_{ij} - \min_j(V_{ij})}{\max_j(V_{ij}) - \min_j(V_{ij})} \tag{6}$$

Table 6

The relevance of the extracted characteristics and the choice of a relevant learning algorithm for the task is a real challenge. These choices have a direct impact on the quality of the system.

Table 3
Details about bike shared dataset.

N	Attribute	Description	Type
1	timestamp	Timestamp field for grouping the data	Numerical
2	cnt	Count of a new bike shares	Numerical
3	t1	Real temperature in °C	Numerical
4	t2	Apparent temperature in °C	Numerical
5	hum	Humidity in percentage (%)	Numerical
6	windspeed	Twind speed in km/h	Numerical
7	weather_code	Category of the weather	Categorical
8	isholiday	Boolean field - 1 holiday / 0 non holiday	Boolean
9	isweekend	Boolean field - 1 if the day is weekend	Boolean
10	season	Category field meteorological seasons	Categorical

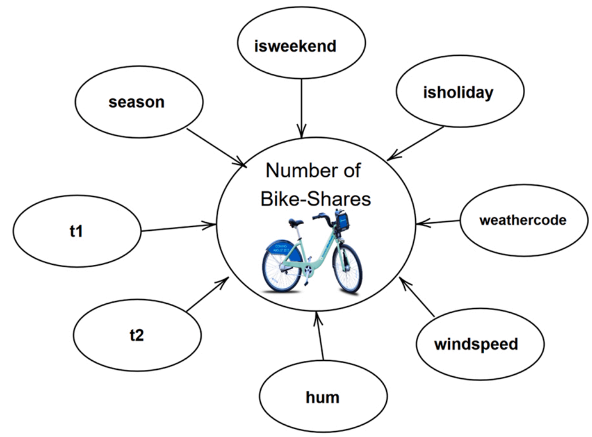


Fig. 6. Parameters that influence the number of bikes share.

5.2. Experimental protocol

The aim of this section is to describe the experimental protocol to test and evaluate the performances of the proposed models for the prediction number of bike shared per hour or day. The implementations have being carried out on a windows platform under python 3.7 using principally sklearn library. The computer used is of the *"Asus"* brand with the following configuration: 8 GB of RAM, intel core i7 processor and an NVIDIA Geforce 930M for graphic card. For the experiments, the dataset has been divided into 5 folds and the tests were performed with cross validation using 4 folds for training and the one for testing.

The performances of the involved algorithms have been compared in terms of r-square (R^2), root mean log square error (RMLSE), mean absolute error (MAE), mean square error (RMSE) and processing time. These performance measure are the commonly used to evaluate de regression models. They have been involved in many previous works associated to bike sharing including Tekouabou et al. (2020).

The MAE defines by Eq. (9) consists of calculating the arithmetic mean of the absolute values of the differences between the actually bike shared rates and those predicted by the machine learning model Tekouabou et al. (2020). Since the objective is to measure the error produced by the BSS, the lower the MAE (or RMSE) obtained, the better the prediction quality, meaning that the predicted performance is closer to the reference. Let $Y = [y_1, \ldots, y_N]$ the predicted bike sharing rates and $\widehat{Y} = [\widehat{y}_1; \ldots; \widehat{y}_N]$ the actual bike shared number for N instances with an average of $\overline{\widehat{y}}$, Eq. (7), (9), (8) are used to compute the RMLSE, MAE, R^2 and RMSE respectively as follows:

$$\text{RMLSE} = \sqrt{\frac{\sum_{i=1}^{N}(\log(y) - \log(\widehat{y}))^2}{N}} \tag{7}$$

$$\text{MSE} = \frac{\sum_{i=1}^{N}(y - \widehat{y})^2}{N} \tag{8}$$

$$\text{MAE} = \frac{\sum_{i=1}^{N}|y - \widehat{y}|}{N} \tag{9}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y - \widehat{y})}{\sum_{i=1}^{N}(\widehat{y} - \overline{\widehat{y}})} \tag{10}$$

The choice of a single metric does not always allow the models to be separated. If the RLMSE and the MSE are known to give the error characterized by the variance and the mean between the predicted and the actual, favouring the effects of large deviations, the absolute error

[1] Https://cycling.data.tfl.gov.uk/
[2] www.freemeteo.com
[3] https://www.gov.uk/bank-holidays
[4] https://www.kaggle.com/hmavrodiev/london-bike-sharing-dataset

may reflect the effect of accuracy in predicting the rate of bike sharing. R^2 will give himself the proportion of the actual rate that was correctly predicted. The optimal model will result from the homogeneity between all three metrics.

5.3. Hyper-parameters tuning and optimization of the models

Optimal models for a specific task are obtained by hyper-parameters adjustments. For each algorithm, we performed several setting for a better prediction of the number of shared bikes. We would like to compare the performance of the different learning models presented in Section 4.3 with the prediction of the number of shared bikes in the city of London. For the bagging methods, the most important hyper-parameters are the number of iterations, the bagging error (E_{OOB}) and the best weak learner. For the boosting method, we had to find the number of iterations, the learning rate and the best weak learner (Tekouabou et al., 2020). Fig. 7 illustrates the tests that were performed to select these parameters using a simple decision tree as a weak learner. Fig. 7a and b provide the optimal parameters for the ABR and XGBR boosting methods, respectively. From Fig. 7a, we can see that optimal for ABR model has a learning rate of 1 while its number of estimators must be greater than 40. From Fig. 7b, we notice that the optimal learning rate and number of estimators for XGBR must be at least greater than 0.05 and 20, respectively. Fig. 7c shows that for the bagging models (BR and RFR), the out-of-bag error (E_{OOB}) is almost equal to 0 for 20 or more estimators. Knowing these best parameters allows us to experiment and find the most efficient model for our prediction task.

5.4. Results analysis

In this section we discuss the obtained results. We first present the data analysis on the total number of bike shared on given BSS of London. We then analyse the predictive model results from our learning algorithms.

5.4.1. Data analysis

Based on our database from the London BSS we described in Section 5.1, we can quickly create diagrams to determine how the number of bike shared is affected by the available features. To do this, we now see some graphs that illustrate these different relations.

According to Fig. 8a, we can segment the day time range into three segments. Thus, after a first peak at the time of arrival at work (7–9 am), the second peak is at lunchtime, between 12 and 1 pm (about 1500 movements per hour). From 5 p.m. to 7 p.m., the number of bike rentals shared is at its highest with nearly 2800 movements per hour, then declines steadily from 2800 to 500. In the evening, the use of the system remains high but gradually decreases, losing 2300 movements per hour between 11 pm and 5 am.

Fig. 8e shows the bike sharing rate obtained as a function of time for the three patterns (workdays, weekend and holidays). For the first pattern ”workdays”, we can notice that there are two peaks of bike-sharing activities from 7:00 am to 9:00 am, then from 5:00 pm to 7:00 pm, i.e. during the periods of going to and from work. For weekend and holiday reasons, we can notice that the activity of shared bikes is from 10 am to 8 pm.

In order to clearly visualise the use of bikes at different times of the day, the bike usage data for each period from 0 to 24 hours is extracted as shown in 8 b. We notice at the beginning of the week the rate of bike sharing is low, in the middle of the week the rate of bike sharing increases. However, during the weekend the number of rentals is much lower.

The use of bikes in different situations is more weather related, so in order to show the impact of weather conditions on the bike usage data for each situation are extracted, as shown in Fig. 8d. This figure shows that the number of bikes shared in better weather conditions is very high compared to bad weather.

Fig. 8c shows the bike sharing rate for each season. It can be noted that the use of shared bikes is quite considerable during summer and fall.

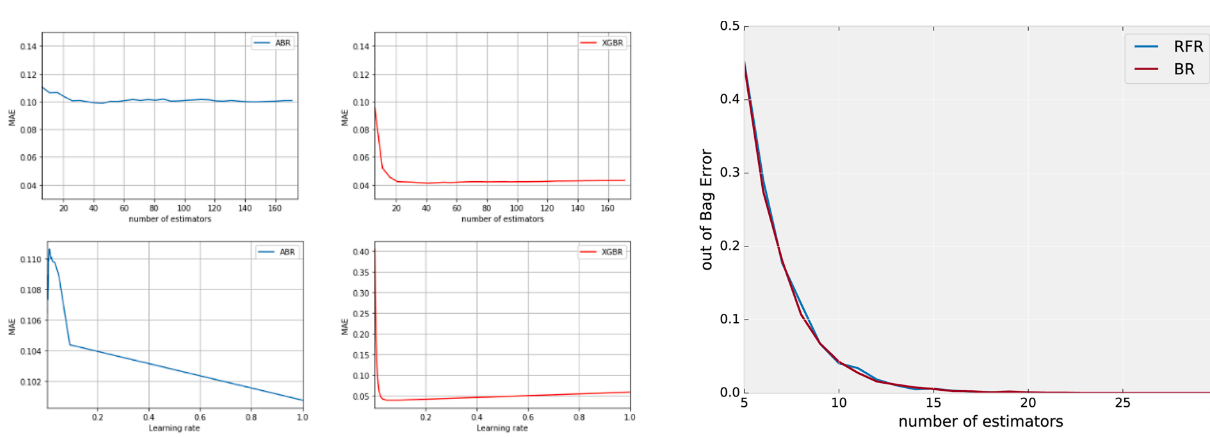
Fig. 8f shows the bike sharing rate obtained as a function of months for the three reasons (workdays, weekends and holidays). For all three reasons, we notice a considerable movement of bike sharing between the months of April and October. For the rest of the year, the sharing rate is very low. This is true for all three patterns.

5.4.2. Predictive result analysis

To show the reliability of our model, we compare for the different algorithms, the validation and test performances by computing the MSE, MAE, RMLSE and R^2 scores of each algorithm. The obtained performance values are grouped below in Table 4 and are also illustrated graphically.

Fig. 9 illustrates the performance of the different algorithms used for the R^2 metric. From Fig. 9, we can see that the bagging methods RFR and BR give the best scores for both validation (0.9532, and 0.9531 respectively) and testing (0.9548 and 0.955 respectively)). They are followed by the boosting algorithms, in particular XGBR then ABR, which give respectively 0.9513 then 0.8264 for validation and 0.9513 then 0.8036 for the test. Finally individual algorithms SVMR then LR give the lowest R^2 score with respectively 0.7892 then 0.6538 for validation and 0.8115 then 0.6807 for the test.

The same trend is confirmed for the other three performance metrics associated with the different calculated errors. Figs. 10–12 illustrate the errors corresponding to the MAE, RMSLE and MSE measurements



(a) (b) (c)

Fig. 7. Tuning and optimization of hyper-parameters of the proposed models.

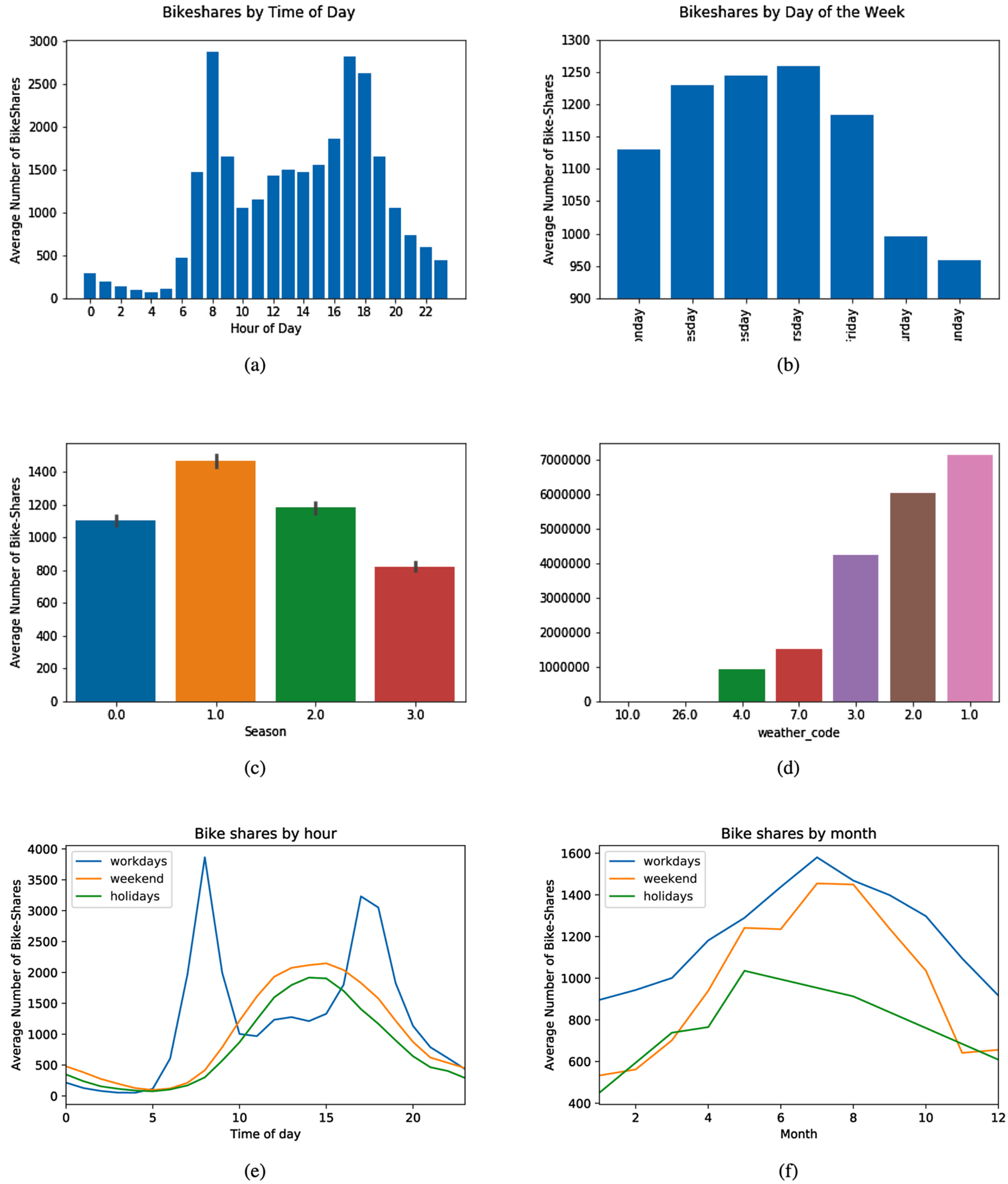


Fig. 8. Distribution of the average number of bike shared: (a) by hour of day; (b) by day of week; (c) by climatic season (d) according to the weather represented by codes, (e) time of day grouped by workdays, weekends and holidays (f) by month-days grouped by workdays, weekends and holidays.

Table 4
Evaluation of BSS performance prediction with ensemble models in terms of MAE, MSE, RMLSE and R^2.

Algorithm	R^2		**MAE**		**MSE**		**RMSLE**	
	Validation score	Test score	Validation score	Test score	Validation score	Test score	Validation score	Test score
LR	0.6538	0.6807	0.1272	0.1241	0.0268	0.0253	0.0821	0.078
RFR	0.9532	0.9548	0.0406	0.0413	0.0036	0.0036	0.0319	0.03
XGBR	0.9513	0.9513	0.0417	0.0421	0.0037	0.0039	0.0348	0.031
SVMR	0.7893	0.8115	0.1057	0.1011	0.0163	0.0149	0.0619	0.0585
ABR	0.8264	0.8036	0.0934	0.101	0.0136	0.0156	0.0596	0.0608
BR	0.9531	0.955	0.0406	0.0412	0.0036	0.0036	0.032	0.0299

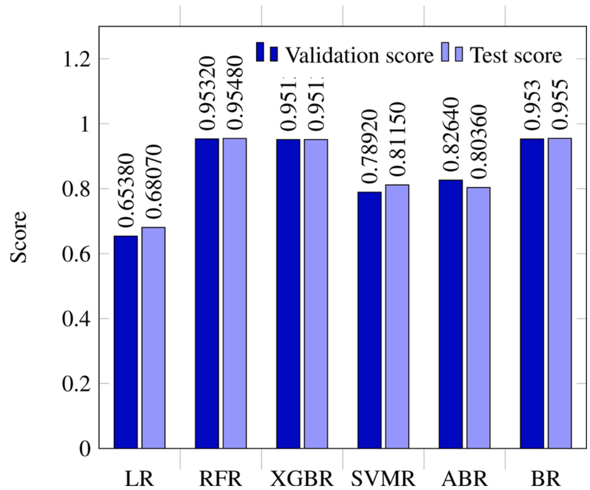


Fig. 9. Comparison of the performances of the used regression algorithms according R^2.

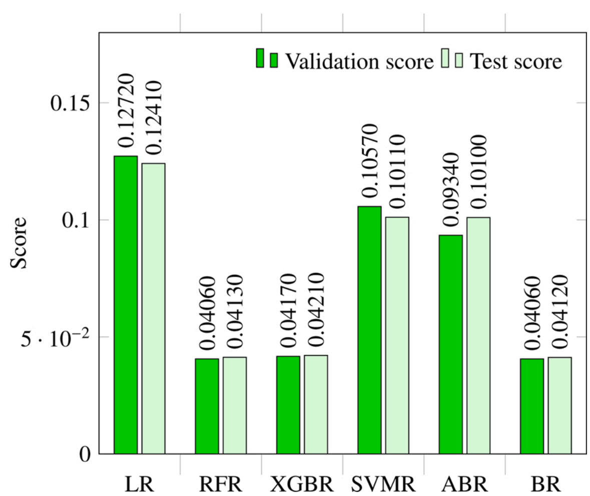


Fig. 10. Comparison of the performances of the used regression algorithms according MAE.

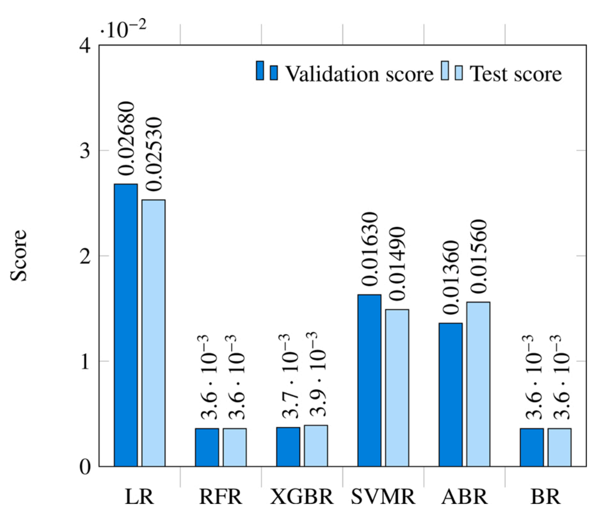


Fig. 11. Comparison of the performances of the used regression algorithms according MSE.

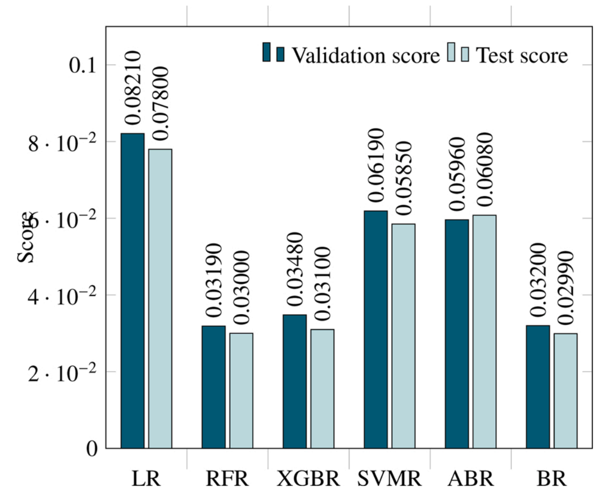


Fig. 12. Comparison of the performances of the used regression algorithms according RMSLE.

respectively. From these figures whose values are summarized in Table 4, we can notice that overall the bagging methods always give the best scores and therefore the lowest errors both in validation and testing. The values obtained for RF and BR are respectively 0.0406, 0.0036 and 0.0319 in validation then 0.0413, 0.0036 and 0.03 in test for MAE, MSE and RMLSE. These best performances are followed by those of boosting methods such as XGBR then ABR in order of performance. Among these two algorithms based on boosting, XGBR is more efficient with respectively 0.0417, 0.0037, 0.0348 in validation then 0.0421, 0.0039, 0.031 in test while ABR gave 0.0934, 0.0136, 0.0596 in validation then 0.101, 0.0156, 0.0608 in test and this, for the trio of metrics MAE, MSE and RMLSE. Finally the individual algorithms that almost gave the lowest score corresponding to the highest errors. For the metrics MAE, MSE and RMLSE, the values obtained are respectively 0.1057, 0.0163, 0.0619 in validation then 0.1011, 0.0149, 0.0585 in test for SVMR and finally the lowest scores of 0.1272, 0.0268, 0.0821 in validation then 0.1241, 0.0253, 0.078 in test attributed to LR for the three metrics. We can see that the bagging technique is more favourable on all metrics to build the best models. The validation score values are roughly equal to the test values, which shows that there is no over-fitting. However, for the individual models, there is some few under-fitting which would have been corrected by the ensemble strategy to give the best scores by eliminating bias and variance.

The conclusions made by performance analysis are also confirmed by the distribution patterns of 200 randomly drawn instances represented in Fig. 13. The points of the distribution correspond to the couple real value of the bikes sharing in abscissa and the predictive values associated in ordinate and this for the six algorithms used. From Fig. 13, we can clearly see that the distribution of RFR and BR is better. Then comes XGBR, then ABR and finally SVMR and LR have the worst distributions.

The next section is devoted to the discussion of the results obtained.

6. Discussion of the results

Smart cities offer a wide variety of ways for citizens to get around depending to their needs. City managers and transportation service providers need to meet both the mobility needs of users, the return on investment and the sustainability (Banerjee, Kabir, Khadem, & Chavis, 2020; Sherriff et al., 2020). These requirements have fostered the emergence of new urban mobility offers such as self-service shared bikes which can be used several times as a sustainable modes of transport responding favourably to the mobility needs of the inhabitants. In urban planning of bike sharing program, the number of bikes shared is not almost known and changes dynamically under the influence of several

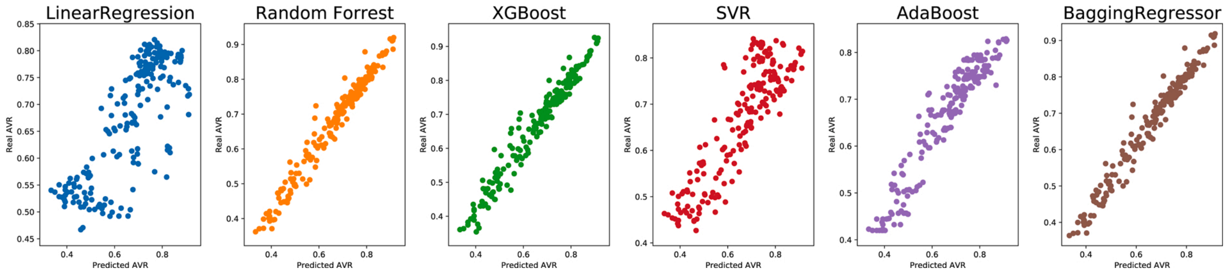


Fig. 13. Comparison of the regression distribution look for 200 first instances by the different algorithms.

factors including spatial distribution of bikes, time, weather conditions, seasons, etc (Ji et al., 2020). The duration of bike use is often unexpected in advance and the number of bikes shared also is often variable. Moreover, all these parameters depend on several factors (Bikeshare, 2013; Sherriff et al., 2020). Collecting the data associated to all these factors through the IoT and analyzing them using machine learning algorithms would help to make BSS more cost-effective and accessible to users. In order to offer more flexibility in service to the user, shared bikes programs can now involves docked and dockless BSS as in London. In this paper, we aim to design a predictive model for self-service bike sharing program planning and case of study in London city. As stated by Chen et al. (2020), self-service bike sharing programs certainly have advantages in terms of flexibility, convenience, accessibility and availability for the user. Even though the majority of BSS in London are docked, there is also another alternative that is increasingly being studied, which is dockless BSS. Indeed, dockless has advantages in terms of flexibility, convenience, accessibility and availability for the user. However, one of the biggest problems of shared bikes is theft and vandalism, which are more increased in self-service shared bike sharing systems (Chen et al., 2020; Floret, 2014; Zhang, Zhang, Duan, & Bryde, 2015). In addition, the business model and operational challenges associated to bike sharing may generate other more urgent problems, such as vandalism and overproduction of bicycles, causing damage to the environment and threatening the sustainability of cities and society (Chen et al., 2020). Forecasting the number of bike shared is therefore very important both for the business model, the sustainability of BSS and therefore of cities. Strategically, planning the number of shared bikes allows to: (1) refine the spatio-temporal distribution of bicycles in the city and therefore automate the display of the number of bicycles available in real time for users as well as for city controllers and promoters (Banerjee et al., 2020; Bikeshare, 2013). (2) For business models, this will allow potential future investors to benchmark and evaluate the profitability in the short, medium and even long term knowing that the turnover is both associated with the number of shared bicycles and the use time that we explore in our model (Bikeshare, 2013; Sherriff et al., 2020). (3) The “Sustainability” already due to the alternative of shared bicycles which is more ecological compared to other modes of urban mobility. The integration of the prediction model not only allows to automate and improve the quality of service offered to users. It also to make the system more intelligent by removing the need for physical resources to attract more users and thus better contribute to the sustainability of the city. For a sustainable city and society, urban mobility is one of the important factors (Sherriff et al., 2020; Sun & Zhai, 2020). The smart mobility solution such as proposed BSP should confirm the three pillars of sustainability including financial sustainability, social sustainability, and environmental sustainability. Financial sustainability can be achieved by providing a low-cost mobility service while improving the accuracy. Social sustainability can be obtained by the customer satisfaction of the shared bike services close to the users and finely tuned to their needs. Environmental sustainability can be achieved by controlling the carbon and other toxic gas emissions by the healthcare equipment. For example, the IoT, mobility sensors, spatial distribution of bikes, station equipment’s if they exist and especially the recycling of used bicycles preserve the environment (Bikeshare, 2013; Chen et al., 2020; Sherriff et al., 2020). Machine learning based models can be applied in real life to manage and plan bike-sharing travels for users over a period of time, as the system can plan the availability of bikes for all users who request a bike-sharing travel. As illustrated in Fig. 3, the training of predictive systems is essentially based on the extraction of characteristics associated with a training algorithm. Data pre-processing steps have made the data more suitable for efficient training of predictive regression models. Several training algorithms were tested including SVMR, LR, BR, RFR, XGBR and ABR. The simulation results show that the RFR, BR, and XGBR achieved the best performances taking into account factors that affect bike usage in BSS. In addition, most literature work does not take into account these factors (especially the meterological factors) that impact bike use in a BSS. (Borgnat et al., 2011; Fanaee-T & Gama, 2014; Kaltenbrunner et al., 2010; Lu et al., 2019; Yoon et al., 2012; Zhang et al., 2019). Our approach has been evaluated by several evaluation metrics (R^2, MAE, MSE and RMSLE) unlike previous works (Borgnat et al., 2011; Fanaee-T & Gama, 2014; Kaltenbrunner et al., 2010; Lu et al., 2019; Yoon et al., 2012; Zhang et al., 2019). Considering factors such as weather in the prediction process guarantees a dynamic aspect of the model we have built compared to other works (Caggiani, Camporeale, Ottomanelli, & Szeto, 2018; Raviv & Kolka, 2013).

The deployment of the proposed model can be done in real time in order to automate and make reliable urgent decisions on the spatial and temporal distribution of bike shared in self-service (Banerjee et al., 2020; Boufidis, Nikiforiadis, Chrysostomou, & Aifadopoulou, 2020; Conrow, Murray, & Fischer, 2018; Winslow & Mont, 2019). This is important to already improve the quality of BSS provided services by allowing automatically both users and providers to have an idea about the number of bicycles available in real time or when needed. Including predictive analysis will thus allow for more autonomous and intelligent management of the BSS. The stakes for investors is to have a clear idea of the profitability of BSS and to be accompanied in making strategic decisions. We complete by a real and concrete observation linked to the fact that today the emergence of BSS depends on the conviction of users to use shared bikes and thus attract investors. This is a major challenge for several countries. In this case, we show that the implementation of an automatic predictive model will improve the service and thus attract the maximum number of users while reassuring investors. The machine learning based forecasting system allows an automatic, dynamic and intelligent parametrization of the BSS. Its stakes for modern BSS are therefore strategic and multidimensional.

The proposed system could thus serve as an effective tool for automating and improving the quality of service providing a dynamic BSS management. However, given all the challenges of management of bike-sharing systems for the companies and governments that oversee these systems, the user decision model and user waiting time could be considered in future work. This will lead our future research to delve deeper into the BSS most compromising between sustainability and user-centric service in the business or strategic model.

7. Conclusion

The new technologies offer by the smart cities and the IoT make daily life more comfortable as was presented at the beginning of this paper. Thanks to these technologies we can retrieve data from several connected sources. These data allow us to better understand how this new form of transportation is used by city inhabitants. Thus, on the basis of real data from three sources of London BSS, we have set out two main objectives in this paper. The first was to study different aspects of the shared bicycle system, through the analysis of travel data taking into account several factors. The second was to develop a prediction approach based on machine learning to help companies and stakeholders in the shared bike sector to manage the system in an efficient and cost-effective way. The simulation results show that the overall methods are more effective than the individual methods. The approach also provides appropriate recommendations for those responsible for the system.

Declaration of Competing Interest

The authors report no declarations of interest.

References

Austwick, M. Z., O'Brien, O., Strano, E., & Viana, M. (2013). The structure of spatial networks and communities in bicycle sharing systems. *PLOS ONE, 8*.

Bachand-Marleau, J., Lee, B. H., & El-Geneidy, A. M. (2012). Better understanding of factors influencing likelihood of using shared bicycle systems and frequency of use. *Transportation Research Record, 2314*, 66–71.

Banerjee, S., Kabir, M. M., Khadem, N. K., & Chavis, C. (2020). Optimal locations for bikeshare stations: A new gis based spatial approach. *Transportation Research Interdisciplinary Perspectives*, 100101.

Beecham, R., & Wood, J. (2014). Exploring gendered cycling behaviours within a large-scale behavioural data-set. *Transportation Planning and Technology, 37*, 83–97.

Bikeshare, C. (2013). *Capital bikeshare member survey report*. Washington, DC.

Bordagaray, M., Ibeas, A., & dell'Olio, L. (2012). Modeling user perception of public bicycle services. *Procedia-Social and Behavioral Sciences, 54*, 1308–1316.

Borgnat, P., Abry, P., Flandrin, P., Robardet, C., Rouquier, J.-B., & Fleury, E. (2011). Shared bicycles in a city: A signal processing and data analysis perspective. *Advances in Complex Systems, 14*, 415–438.

Borgnat, P., Robardet, C., Abry, P., Flandrin, P., Rouquier, J.-B., & Tremblay, N. (2013). A dynamical network view of lyon's vélo'v shared bicycle system. *in: Dynamics On and Of Complex Networks, Volume 2*, 267–284.

Borgnat, P., Fleury, E., Robardet, C., & Scherrer, A. Spatial analysis of dynamic movements of vélo'v, lyon's shared bicycle program.

Bortolussi, L., & Hillston, J. (2015). *Efficient checking of individual rewards properties in markov population models*. arXiv preprint arXiv:1509.08561.

Boufidis, N., Nikiforiadis, A., Chrysostomou, K., & Aifadopoulou, G. (2020). Development of a station-level demand prediction and visualization tool to support bike-sharing systems' operators. *Transportation Research Procedia, 47*, 51–58.

Bouveyron, C., Côme, E., Jacques, J., et al. (2015). The discriminative functional mixture model for a comparative analysis of bike sharing systems. *The Annals of Applied Statistics, 9*, 1726–1760.

Bouzguenda, I., Alalouch, C., & Fava, N. (2019). Towards smart sustainable cities: A review of the role digital citizen participation could play in advancing social sustainability. *Sustainable Cities and Society*, 101627.

Breiman, L., & Cutler, A. (2005). *Random forests. berkeley*.

Breiman, L. (1996). Bagging predictors. *Machine learning, 24*, 123–140.

Buck, D., Buehler, R., Happ, P., Rawls, B., Chung, P., & Borecki, N. (2013). Are bikeshare users different from regular cyclists? a first look at short-term users, annual members, and area cyclists in the Washington, DC, region. *Transportation research record, 2387*, 112–119.

Bullock, C., Brereton, F., & Bailey, S. (2017). The economic contribution of public bike-share to the sustainability and efficient functioning of cities. *Sustainable cities and society, 28*, 76–87.

Caggiani, L., Camporeale, R., Ottomanelli, M., & Szeto, W. Y. (2018). A modeling framework for the dynamic management of free-floating bike-sharing systems. *Transportation Research Part C: Emerging Technologies, 87*, 159–182.

Cao, Y., & Shen, D. (2019). Contribution of shared bikes to carbon dioxide emission reduction and the economy in Beijing. *Sustainable Cities and Society, 51*, 101749.

Castillo-Manzano, J. I., Castro-Nuno, M., & Lopez-Valpuesta, L. (2015). Analyzing the transition from a public bicycle system to bicycle ownership: A complex relationship. *Transportation Research Part D: Transport and Environment, 38*, 15–26.

Chen, T. & Guestrin, C. Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining 785-794.

Chen, T., & Ren, J. (2009). Bagging for Gaussian process regression. *Neurocomputing, 72*, 1605–1610.

Chen, Q., & Sun, T. (2015). A model for the layout of bike stations in public bike-sharing systems. *Journal of Advanced Transportation, 49*, 884–900.

Chen, M., Miao, Y., Hao, Y., & Hwang, K. (2017). Narrow band internet of things. *IEEE Access, 5*, 20557–20577.

Chen, Z., van Lierop, D., & Ettema, D. (2020). Dockless bike-sharing systems: what are the implications? *Transport Reviews*, 1–21.

Conrow, L., Murray, A. T., & Fischer, H. A. (2018). An optimization approach for equitable bicycle share station siting. *Journal of Transport Geography, 69*, 163–170.

Coppi, R., D'Urso, P., Giordani, P., & Santoro, A. (2006). Least squares estimation of a linear regression model with lr fuzzy response. *Computational Statistics & Data Analysis, 51*, 267–286.

Cutler, A., Cutler, D. R., & Stevens, J. R. (2012). Random forests. *in: Ensemble machine learning*, 157–175.

Dell'Olio, L., Ibeas, A., & Moura, J. L. Implementing bike-sharing systems, in: Proceedings of the Institution of Civil Engineers-Municipal Engineer, volume 164, Thomas Telford Ltd, 89-101.

DeMaio, P. (2009). Bike-sharing: History, impacts, models of provision, and future. *Journal of Public Transportation, 12*, 3.

Doorley, R., Pakrashi, V., & Ghosh, B. (2015). Quantifying the health impacts of active travel: assessment of methodologies. *Transport Reviews, 35*, 559–582.

Eren, E., & Uz, V. E. (2019). *A review on bike-sharing: The factors affecting bike-sharing demand* (p. 101882). Sustainable Cities and Society.

Etienne, C., & Latifa, O. (2014). Model-based count series clustering for bike sharing system usage mining: a case study with the vélib'system of paris. *ACM Transactions on Intelligent Systems and Technology (TIST), 5*, 1–21.

Faghih-Imani, A., Eluru, N., El-Geneidy, A. M., Rabbat, M., & Haq, U. (2014). How land-use and urban form impact bicycle flows: evidence from the bicycle-sharing system (bixi) in montreal. *Journal of Transport Geography, 41*, 306–314.

Fanaee-T, H., & Gama, J. (2014). Event labeling combining ensemble detectors and background knowledge. *Progress in Artificial Intelligence, 2*, 113–127.

Fishman, E., Washington, S., & Haworth, N. L. (2012). *An evaluation framework for assessing the impact of public bicycle share schemes*.

Fishman, E., Washington, S., & Haworth, N. (2013). Bike share: a synthesis of the literature. *Transport Reviews, 33*, 148–165.

Fishman, E. (2016). Bikeshare: A review of recent literature. *Transport Reviews, 36*, 92–113.

Floret, M. (2014). *Bike-sharing and the promotion of urban sustainable mobility, Ph.D. thesis, uniwien*.

Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences, 55*, 119–139.

Freund, Y., Schapire, R. E., et al. Experiments with a new boosting algorithm, in: icml, volume 96, Citeseer 148-156.

Freund, Y., Schapire, R., & Abe, N. (1999). A short introduction to boosting. *Journal-Japanese Society For Artificial Intelligence, 14*, 1612.

Froehlich, J., Neumann, J., & Oliver, N. (2008). Measuring the pulse of the city through shared bicycle programs. *Proc. of UrbanSense, 08*, 16–20.

Froehlich, J. E., Neumann, J., & Oliver, N. Sensing and predicting the pulse of the city through shared bicycling, in: Twenty-First International Joint Conference on Artificial Intelligence.

Fuller, D., Gauvin, L., Kestens, Y., Daniel, M., Fournier, M., Morency, P., & Drouin, L. (2011). Use of a new public bicycle share program in montreal, canada. *American Journal of Preventive Medicine, 41*, 80–83.

Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media.

García-Palomares, J. C., Gutiérrez, J., & Latorre, M. (2012). Optimizing the location of stations in bike-sharing programs: A gis approach. *Applied Geography, 35*, 235–246.

Garcia-Gutierrez, J., Romero-Torres, J., & Gaytan-Iniestra, J. (2014). Dimensioning of a bike sharing system (bss): a study case in Nezahualcoyotl, Mexico. *Procedia-Social and Behavioral Sciences, 162*, 253–262.

Hoe, N., & Kaloustian, T. (2014). *Bike sharing in low-income communities: An analysis of focus groups findings fall 2014*.

Jcdecaux developer https://developer.jcdecaux.com/.

Jensen, P., Rouquier, J.-B., Ovtracht, N., & Robardet, C. (2010). Characterizing the speed and paths of shared bicycle use in lyon. *Transportation Research part D: Transport and Environment, 15*, 522–524.

Ji, Y., Ma, X., He, M., Jin, Y., & Yuan, Y. (2020). Comparison of usage regularity and its determinants between docked and dockless bike-sharing systems: A case study in Nanjing, China. *Journal of Cleaner Production, 255*, 120110.

Kaltenbrunner, A., Meza, R., Grivolla, J., Codina, J., & Banchs, R. (2010). Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system. *Pervasive and Mobile Computing, 6*, 455–466.

Kim, D., Shin, H., Im, H., & Park, J. Factors influencing travel behaviors in bikesharing, in: Transportation Research Board 91st Annual Meeting.

Krykewycz, G. R., Puchalsky, C. M., Rocks, J., Bonnette, B., & Jaskiewicz, F. (2010). Defining a primary market and estimating demand for major bicycle-sharing program in Philadelphia, Pennsylvania. *Transportation Research Record, 2143*, 117–124.

Lathia, N., Ahmed, S., & Capra, L. (2012). Measuring the impact of opening the London shared bicycle scheme to casual users. *Transportation Research part C: Emerging Technologies, 22*, 88–102.

Lazarus, J., Pourquier, J. C., Feng, F., Hammel, H., & Shaheen, S. (2020). Micromobility evolution and expansion: Understanding how docked and dockless bikesharing models complement and compete—a case study of San Francisco. *Journal of Transport Geography, 84*, 102620.

Leister, E. H., Vairo, N., Sims, D., & Bopp, M. (2018). Understanding bike share reach, use, access and function: An exploratory study. *Sustainable Cities and Society, 43*, 191–196.
Lin, J.-R., & Yang, T.-H. (2011). Strategic design of public bicycle sharing systems with service level constraints. *Transportation Research part E: Logistics and Transportation Review, 47*, 284–294.
Lohry, G. F. & Yiu, A. Bikeshare in China as a public service: Comparing government-run and public-private partnership operation models, in: Natural Resources Forum, volume 39, Wiley Online Library, 41-52.
London, T. (2013). *Travel in London*. London, UK: Transport London.
Lu, M., An, K., Hsu, S.-C., & Zhu, R. (2019). Considering user behavior in free-floating bike sharing system design: A data-informed spatial agent-based model. *Sustainable Cities and Society, 49*, 101567.
Mattson, J., & Godavarthy, R. (2017). Bike share in fargo, north dakota: Keys to success and factors affecting ridership. *Sustainable Cities and Society, 34*, 174–182.
Mavroforakis, M. E., & Theodoridis, S. (2006). A geometric approach to support vector machine (svm) classification. *IEEE Transactions on Neural Networks, 17*, 671–682.
Midgley, P. (2009). The role of smart bike-sharing systems in urban mobility. *Journeys, 2*, 23–31.
Midgley, P. (2011). Bicycle-sharing schemes: enhancing sustainable mobility in urban areas, United Nations. *Department of Economic and Social Affairs, 8*, 1–12.
Mishra, S., Mishra, D., & Santra, G. H. (2017). Adaptive boosting of weak regressors for forecasting of crop production considering climatic variability: An empirical assessment. *Journal of King Saud University-Computer and Information Sciences*.
Morency, C., Trepanier, M., & Godefroy, F. (2011). *Insight into Montreal's Bikesharing System, Technical Report*.
Murphy, H. (2010). *Dublin bikes: An investigation in the context of multimodal transport*. Dublin: MSc Sustainable Development, Dublin Institute of Technology.
Nair, R., & Miller-Hooks, E. (2011). Fleet management for vehicle sharing operations. *Transportation Science, 45*, 524–540.
Nair, R., Miller-Hooks, E., Hampshire, R. C., & Bušić, A. (2013). Large-scale vehicle sharing systems: analysis of vélib'. *International Journal of Sustainable Transportation, 7*, 85–106.
Noland, R. B., & Ishaque, M. M. (2006). Smart bicycles in an urban area: Evaluation of a pilot scheme in London. *Journal of Public Transportation, 9*, 5.
O'brien, O., Cheshire, J., & Batty, M. (2014). Mining bicycle sharing data for generating insights into sustainable transport systems. *Journal of Transport Geography, 34*, 262–273.
Ogilvie, F., & Goodman, A. (2012). Inequalities in usage of a public bicycle sharing scheme: socio-demographic predictors of uptake and usage of the London (uk) cycle hire scheme. *Preventive Medicine, 55*, 40–45.
Parkes, S. D., Marsden, G., Shaheen, S. A., & Cohen, A. P. (2013). Understanding the diffusion of public bikesharing systems: evidence from europe and north america. *Journal of Transport Geography, 31*, 94–103.
Randriamanamihaga, A. N., Côme, E., Oukhellou, L., & Govaert, G. (2014). Clustering the vélib? dynamic origin/destination flows using a family of poisson mixture models. *Neurocomputing, 141*, 124–138.
Raviv, T., & Kolka, O. (2013). Optimal inventory management of a bike-sharing station. *IIE Transactions, 45*, 1077–1093.
Razi, M. A., & Athappilly, K. (2005). A comparative predictive analysis of neural networks (nns), nonlinear regression and classification and regression tree (cart) models. *Expert Systems with Applications, 29*, 65–74.
Rifkin, J. (2011). *The third industrial revolution: how lateral power is transforming energy, the economy, and the world*. Macmillan.
Rojas-Rueda, D., de Nazelle, A., Tainio, M., & Nieuwenhuijsen, M. J. (2011). The health risks and benefits of cycling in urban environments compared with car use: health impact assessment study. *BMJ, 343*, d4521.
Saharidis, G., Fragkogios, A., & Zygouri, E. A multi-periodic optimization modeling approach for the establishment of a bike sharing network: a case study of the city of athens, in: Proceedings of the International MultiConference of Engineers and Computer Scientists, volume 2, 1226-1231.
Sarkar, A., Lathia, N., & Mascolo, C. (2015). Comparing cities' cycling patterns using online shared bicycle maps. *Transportation, 42*, 541–559.
Sayarshad, H., Tavassoli, S., & Zhao, F. (2012). A multi-periodic optimization formulation for bike planning and bike utilization. *Applied Mathematical Modelling, 36*, 4944–4951.
Schuijbroek, J., Hampshire, R. C., & Van Hoeve, W.-J. (2017). Inventory rebalancing and vehicle routing in bike sharing systems. *European Journal of Operational Research, 257*, 992–1004.
Shaheen, S. A., Guzman, S., & Zhang, H. (2010). Bikesharing in Europe, the Americas, and Asia: Past, present, and future. *Transportation Research Record, 2143*, 159–167.
Shaheen, S. A., Zhang, H., Martin, E., & Guzman, S. (2011). China's hangzhou public bicycle: understanding early adoption and behavioral response to bikesharing. *Transportation Research Record, 2247*, 33–41.
Sharifi, A. (2020). A typology of smart city assessment tools and indicator sets. *Sustainable Cities and Society, 53*, 101936.
Sherriff, G., Adams, M., Blazejewski, L., Davies, N., & Kamerāde, D. (2020). From mobike to no bike in greater Manchester: Using the capabilities approach to explore Europe's first wave of dockless bike share. *Journal of Transport Geography, 86*, 102744.
Sinha, R. S., Wei, Y., & Hwang, S.-H. (2017). A survey on lpwa technology: Lora and nb-iot. *Ict Express, 3*, 14–21.
Subramaniyan, A. B. & Pan, R. Model reuse with bike rental station data, in: DC@ PKDD/ECML.
Sun, C., & Zhai, Z. (2020). The efficacy of social distance and ventilation effectiveness in preventing covid-19 transmission. *Sustainable Cities and Society, 62*, 102390.
Tay, W.-L., Chui, C.-K., Ong, S.-H., & Ng, A. C.-M. (2013). Ensemble-based regression analysis of multimodal medical data for osteopenia diagnosis. *Expert Systems with Applications, 40*, 811–819.
Tekouabou, S. C. K., Cherif, W., Silkan, H., et al. (2020). Improving parking availability prediction in smart cities with iot and ensemble-based model. *Journal of King Saud University-Computer and Information Sciences*.
Transportation, I. for N. Development Policy (New York, A. Gauthier, The bike-share planning guide, ITDP Institute for Planning & Development Policy, 2013.
Tribastone, M., Clark, A., Gast, N., Gilmore, S., & Reijsbergen, D. (2014). Data validation and requirements for case studies. *QUANTICOL Deliverable, 5*.
Tuama, D.Ó. (2015). Ripples through the city: Understanding the processes set in motion through embedding a public bike sharing scheme in a city. *Research in Transportation Business & Management, 15*, 15–27.
Ueasangkomsate, P. Efficiency management of public bike-sharing system in Bangkok.
Vogel, P. & Mattfeld, D. C. Modeling of repositioning activities in bike-sharing systems, in: World conference on transport research (WCTR).
Vogel, P., Saavedra, B. A. N., & Mattfeld, D. C. A hybrid metaheuristic to solve the resource allocation problem in bike sharing systems, in: International Workshop on Hybrid Metaheuristics, Springer, 16-29.
Vogel, P., Greiser, T., & Mattfeld, D. C. (2011). Understanding bike-sharing systems using data mining: Exploring activity patterns. *Procedia-Social and Behavioral Sciences, 20*, 514–523.
Waserhole, A., Jost, V., & Brauner, N. (2013). Pricing techniques for self regulation in vehicle sharing systems. *Electronic Notes in Discrete Mathematics, 41*, 149–156.
Winslow, J., & Mont, O. (2019). Bicycle sharing: Sustainable value creation and institutionalisation strategies in Barcelona. *Sustainability, 11*, 728.
Woodcock, J., Tainio, M., Cheshire, J., O'Brien, O., & Goodman, A. (2014). Health effects of the London bicycle sharing system: health impact modelling study. *BMJ, 348*, g425.
Xu, H., Ying, J., Wu, H., & Lin, F. (2013). Public bicycle traffic flow prediction based on a hybrid model. *Applied Mathematics & Information Sciences, 7*, 667.
Yang, T., Haixiao, P., & Qing, S. Bike-sharing systems in beijing, shanghai and hangzhou and their impact on travel behaviour, in: Transportation Research Board Annual Meeting.
Yigitcanlar, T., Kamruzzaman, M., Foth, M., Sabatini-Marques, J., da Costa, E., & Ioppolo, G. (2019). Can cities become smart without being sustainable? a systematic review of the literature. *Sustainable Cities and Society, 45*, 348–365.
Yoon, J. W., Pinelli, F., & Calabrese, F. (2012). Cityride: a predictive bike sharing journey advisor. *in: 2012 IEEE 13th international conference on mobile data management, IEEE*, 306–311.
Zguira, Y., Rivano, H., & Meddeb, A. (2018). Internet of bikes: A dtn protocol with data aggregation for urban data collection. *Sensors, 18*, 2819.
Zhang, L., Zhang, J., Duan, Z.-y., & Bryde, D. (2015). Sustainable bike-sharing systems: characteristics and commonalities across cases in urban china. *Journal of Cleaner Production, 97*, 124–133.
Zhang, D., Qian, L., Mao, B., Huang, C., Huang, B., & Si, Y. (2018). A data-driven design for fault detection of wind turbines using random forests and xgboost. *IEEE Access, 6*, 21020–21031.
Zhang, Y., Wen, H., Qiu, F., Wang, Z., & Abbas, H. (2019). Ibike: Intelligent public bicycle services assisted by data analytics. *Future Generation Computer Systems, 95*, 187–197.