



Parking Availability Prediction in Smart City

El Arbi Abdellaoui Alaoui^{1,2(✉)} and
Stephane Cedric Koumetio Tekouabou^{3,4} 

- ¹ EIGSI-Casablanca, 282 Route of the Oasis, Casablanca, Morocco
abdellaoui.e@gmail.com
- ² Department Computer Science, Faculty of Sciences and Technologies-My Ismail
University, Errachidia, Morocco
- ³ Laboratory LAROSERI, Department of Computer Science, Faculty of Sciences,
Chouaib Doukkali University, B.P. 20, 24000 El Jadida, Morocco
ctekouaboukoumetio@gmail.com
- ⁴ Center of Urban Systems (CUS), Mohamed VI Polytechnic University (UM6P),
Hay Moulay Rachid, 43150 Ben Guerir, Morocco

Abstract. Smart cities are part of continuous advances in technology to provide a better quality of life for its inhabitants. Important topics in smart cities include urban mobility. Due to the increasing number of vehicles passing through cities, traffic and traffic are becoming common problems. In addition, finding places to park in the city center is not a simple task for drivers. In fact, drivers looking for parking spaces contribute up to 30% traffic congestion. In this context, this paper proposes a model for predicting the availability of parking spaces in car parks using different data sources, regression techniques and machine learning. The objective is to propose a model capable of making an enlightened estimate of the number of free parking places.

Keywords: Smart parking systems · Internet of Things · Smart city · Machine learning

1 Introduction

The continuous growth of the population, the gradual migration of the population to the cities and the progress of information and communication technologies (ICT) have given rise to the phenomenon of “smart cities” [1]. In particular, smart cities are a growing global trend, which aims to integrate ICT solutions to improve the quality of life of its citizens and their interaction with government officials. Considering that traffic and urban mobility are one of the major problems of city development, they face the challenges of sustainable mobility in limited physical spaces. This growing demand is normally limited by the limited physical capacity of a city’s transportation system, traffic and parking [3, 11].

The urban mobility refers to the management of the means of transportation in the city and the costs, time and money, which force citizens to move from

one place to another to carry out their daily activities [11]. A smart city must engage in efficient and multimodal use of transport, promoting public transport and options with less impact on environmental pollution. In large cities, a large number of citizens use public systems (for example, collectives) and private companies (eg. motorists), and many of them depend on how well it works. These types of questions, almost routine, of people affected by the experience they perceive of transportation services are key factors when considering alternatives to improve it.

The urban mobility is one of the typical applications of a smart city, for example, via public transport applications or the provision of customized routines to a user. For the design of these applications (usually with mobile device support), valuable information must be provided to users to enable intelligent travel choices [5]. At the same time, transport companies must commit to improving the quality of the services provided, which also depends on the quantity and quality of information provided by users.

With regard to traffic, the traffic flow makes traffic and parking in various areas of an intermediate city, such as the downtown area, tedious. Currently, some tools can help drivers by reporting traffic flows to avoid crowded areas or to report accidents that do not allow normal operation of streets, roads or highways [2, 6]. However, one of the aspects that affects driver activity is the search for parking spaces. This activity not only affects drivers who want to park, but also the fluidity of traffic in the city. For example, according to [4], 30% of traffic jams are caused by vehicles in search of parking. In this context, if drivers could know in advance which parking spaces were available, it would make it easier for them and would also help to manage the flow of traffic in a more orderly manner.

In this work, a data-centric approach is proposed that aims to predict the percentage of availability of parking spaces for a given city parking. This approach uses machine learning techniques and also integrates different sources of information. The main sources of data used for the development of our model is the data of the car parks of the city Birmingham. For the predictions, different regression techniques were used that predict a block occupancy numeric value for a given date and time, and interesting results were obtained with predictors based on different regression techniques.

1.1 Main Contributions

The key contributions of this work are :

- We show interest in using the Smart City, IoT and ITS concept in public transport;
- We will propose an efficient architecture for smart parking;
- We will detail a model for predicting the availability of places in a smart parking;
- We will use **new** machine learning techniques to predict the availability of places in a smart parking;
- We will evaluate the proposed model using performance metrics.

The rest of this paper is organized as following: Sect. 2 presents the preliminary concepts we will need. Section 3 presents the architecture of our system. In Sect. 4, we presents our methodology for prediction of the availability of parking. Section 5 presents the Experimentation Results and discussion. We conclude the paper in Sect. 6.

2 Smart Cities and Parking Prediction Challenges

In the literature we can find many definitions of a “smart city” [7,9]. The definition we propose is the one taken from [10] because it seems to us to be the most appropriate for our conception of such a system. The authors define a “smart city” as an instrumented, interconnected and intelligent city. Instrumentation enables the capture and integration of real-world data through sensors, counters, vending machines, personal digital devices, image acquisition systems, smart phones, implanted medical devices, the web and other data acquisition systems, including social networks such as human sensor networks. Interconnection means the integration of this data into a company’s IT platform and the communication of this information among various services. Intelligence refers to the presence of complex analyses, modeling, optimization and visualization in operational business processes to make the best operational decisions (e. g. using machine learning).

Many concepts are linked to the concept of “smart city”. These concepts have varying definitions and have both complicated and independent uses. They can still be classified according to three dimensions [8]: technological, human and institutional. These three dimensions are mutually connected to each other to give the concept of “smart city” (Fig. 3).

The technological dimension focuses on infrastructure and software, the human dimension on creativity, diversity and education, and the institutional dimension on politics and governance. With regard to the technological dimension, a properly functioning infrastructure is absolutely necessary but not sufficient to become a “smart city”. An ICT-based structure is a prerequisite, but without a real commitment and willingness to collaborate between public institutions, the private sector, associations, educational institutions and citizens, there would be no “smart city”.

Generally a smart city is based on the following three streams:

- Logistics flow: Logistics flow in a “smart city” corresponds to urban traffic systems (URBAN Traffic Systems -UTS).
- Energy flows: Urban energy flows correspond to all energy transfers between production sources on the one hand and storage systems and/or loads (habitat, public lighting, charging stations, etc.) on the other piece.
- Data flow: With the arrival of new technologies in cities (smart phones, sensor networks, demotic and immotic systems,. . .), a lot of data from different applications is stored and can be transformed into Knowledge.

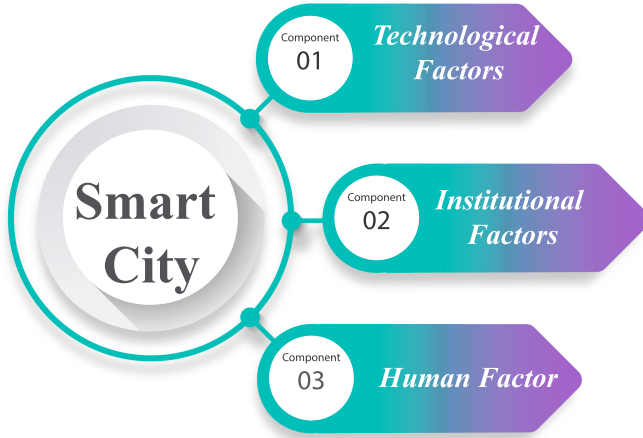


Fig. 1. Components of smart city

A smart city can be improved with the mechanisms of IoT and STIs. Indeed, the concepts of IoT and STI can offer valuable real-time information to smart city players. For example, the various services provided by ITS in a smart city are shown in Figure X: Smart car parks are a privileged application area for ITS and IoT, due to the need for good articulation between many players in the exchange of information. Indeed, THE ITS and IoT are a key to implementing efficient modes of public transport as well as to respond to the problems of timing and quality demanded by customers (Fig. 1).

2.1 Intelligent Transport System (ITS)

The public transport in the city is a privileged field of application of intelligent transport system (ITS), because of the need for good articulation between many actors in the exchange of information. Indeed, the intelligent transport system (ITS) is a central key to implement efficient common transport modes as well as to answer the issues of timing and quality demanded by customers [12].

Intelligent transport systems (ITS) use new information and communication technologies (NTICs) to make transport more automated, thereby increasing the performance of these systems beyond the limits sensory impairment of human drivers. An ITS has the potential to increase traffic safety, its efficiency, driving comfort and reduce the negative impact of transport on the environment. With these different benefits, the development of ITS is actively pushed by Governments, automobile manufacturers and global regulators and standardization bodies [12]. An example of ITS application is the alert of road users to a critical event on the track (accident, breakdown, unworkable pavement, track work, etc.). Other examples are electronic tolls, traffic information systems, and navigation systems. In all these cases, the most important technology to meet these needs is wireless communication. This allows road users to share information

and manage their behavior cooperatively. The various services provided by the TSIs are illustrated in Fig. 2:

- Mobility assistance: Concerning modal choice assistance, the new ITS helps users to determine the appropriate mode of transport according to their needs (speed, cost, safety...). In the route selection help pane, they offer routes that match user expectations. For the assistance of the choice of schedules, they inform the users of the travel schedules for example (peak hours). Booking systems also simplify the management of reservations in real time and remotely (car rental, bus, parking...). Concerning fleet management, the ITS allows controlling the management of products according to the logistics of the company.
- Real-time travel assistance: Electronic payment of ticketing, e-tolls, toll-free, parking or other payment allows people to adjust their costs, save time substantially, etc. Real-time travel help and advising is a real-time traffic information service (their probable journey time, etc.). The monitoring of the fleets allows to control access to certain sensitive or forbidden areas and to increase the safety and security of the transport of dangerous goods, etc.
- Road safety assistance: Driving assistance is used to improve the safety of users and to ensure the comfort of people. These applications include the automatic gearbox, programmable electronic stabilizer, speed limiter, anti-collision system, GPS navigation assistance, etc. For the knowledge of the regulations: geolocation systems used to know the regulations in urban areas, etc. Regarding the application of the regulations, we mention for example radars, the use of cameras in public transport to fight against fraud, etc.

2.2 Internet of Things (IoT)

The Internet of things (IoT) has the potential to transform the transportation industry by profoundly modifying how to gather data, connect with users and automate processes. IoT consists of networking physical objects that, through the use of embedded sensors, actuators, and other devices, can collect and transmit information about network activity in real time. The data collected from this equipment can then be analyzed by the transport agencies for:

- * Improve passenger experience with more reliable transportation, better customer service, more accurate communication and information.
- * Increase safety, including the operation of the transport system through data sensors that detect all anomalies in train speed, pavement temperatures, the condition of aircraft parts, or the number of cars waiting at an intersection.
- * Reduce energy use and congestion through real-time data mining to facilitate the adaptation of resources to operators who need to meet demand, through agility to react quickly to traffic patterns in change, or to address the impact of traffic on fuel consumption, the environment and regional economic competitiveness.

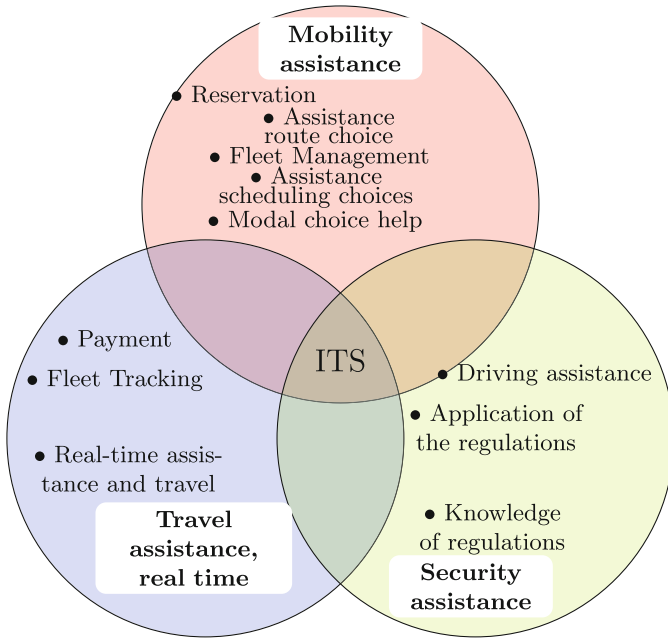


Fig. 2. Different services provided by ITS

3 System Model of Smart Parking

The Intelligent Parking System is an intelligent parking system that uses a detection device to define the occupancy rate of the parking space. It helps the driver to park safely and informs him/her of the availability of parking spaces through appropriate vehicle management. Thanks to intelligent technologies, optimized parking can reach the city centre. A sensor system indicating to drivers where the nearest free parking space is located has already been successfully tested in multi-storey car parks. Street tests are currently underway. In San Francisco, CA, 6,000 sensors have been embedded in the asphalt and are working in conjunction with an application and a GPS. An advanced smart parking system architecture needs to have the following elements:

- Sensor: In each and every parking facility, a set of RF sensors have to be installed. These sensors would relay real-time data about the availability (or otherwise) of slots to nearby drivers. These ‘occupancy sensors’ need to run on battery (wire-connectivity is impractical, given the high number of sensors that have to be used), have long-range RF capabilities (either with a mesh network or with LoRa technology), and offer excellent accuracy. The number of ‘false positives’ generated by a sensor have to be minimal.
- Gateway hardware: Changes in the availability of parking slots will be reflected through a status change of the sensors – and that, in turn, will be



Fig. 3. Smart parking system

collected in a gateway. Unlike the sensors, the gateways have to be operational round-the-clock (the sensors are functional only when their status changes). The data collected in the gateway are then sent on to the centralized server.

- Server: The server in a smart parking setup needs to have 2-way communication protocol support. On the one hand, it will receive the data from the gateways – to generate updated information about the free and occupied parking slots in any facility. At the other end, it will send real-time notifications to the dedicated mobile app in the user’s device – to guide them to the parking area, and the particular ‘vacant’ parking slot. A detailed map of the parking facility will also be sent from the server to the user, for additional guidance.
- Mobile application: The final component of the smart parking architecture is the dedicated mobile app – which serves as the touchpoint for the final users (the app can be installed on smartphones and tablets). After receiving notifications from the server, the app would help the driver to navigate to the empty parking slot (no more confused driving around and resultant frustrations!). For this navigation, the application will either rely on the tools placed in the parking slots (off-road parking) or the phone GPS (for on-road parking).

4 Methodology

The prediction process is shown in Fig. 4 and consisted of several phases including data collection and preprocessing, application of machine learning models, evaluation and prediction.

4.1 About Ensemble-Based Models

Ensemble-based prediction methods combine several independent basic models that are in most cases decision trees or neural networks. Each of these basic models provides an alternative prediction of the problem and the final prediction is a combination (usually by weighted or unweighted vote) of alternative predictions.

The prediction technique by combining the predictions of a set of individual base models generally allows for more stable and accurate output prediction because the error is much smaller than that provided by one of the individual base models which form the overall model. Indeed, the final ensemble-based model corrects the errors made individually by the basic models so that to drastically reduce the total error. To be so effective, the basic models should be forced to fulfill two conditions namely to be independent and to be weak models.

The initial idea was to divide the training data D into n basic data to train n models m_1, m_2, \dots, m_n . But this technique was quickly exceeded because promotes underfitting when n becomes high. To overcome this limit, the methods of resampling the training data into n independent and larger data subsamples to generate weak models. To do this, various techniques among which the most used known are: bagging and boosting.

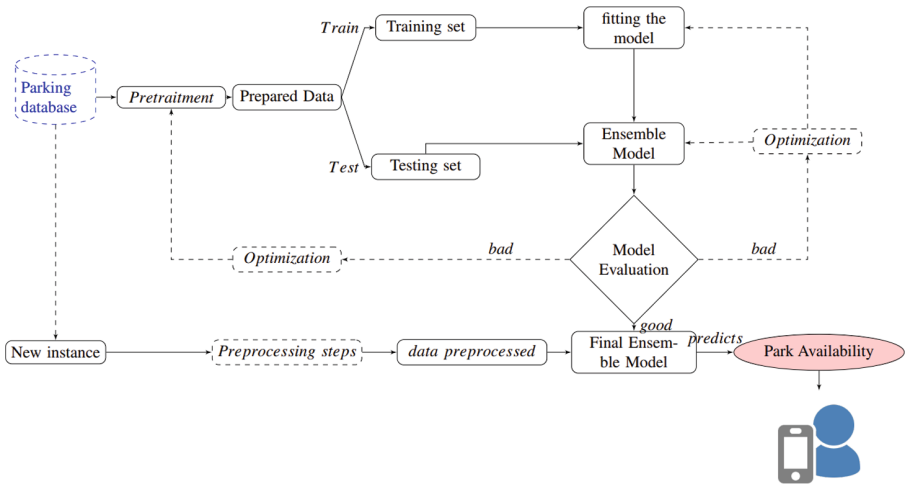


Fig. 4. Global ensemble-based System for Real-time Parking availability Prediction

4.2 Ensemble-Based Models for Regression

The general idea of the ensemble methods is summarized in Fig. 5 which shows that these models are based on three main stages namely: bootstrapping, intermediate modeling and aggregation. The bootstrapping consists in dividing less the

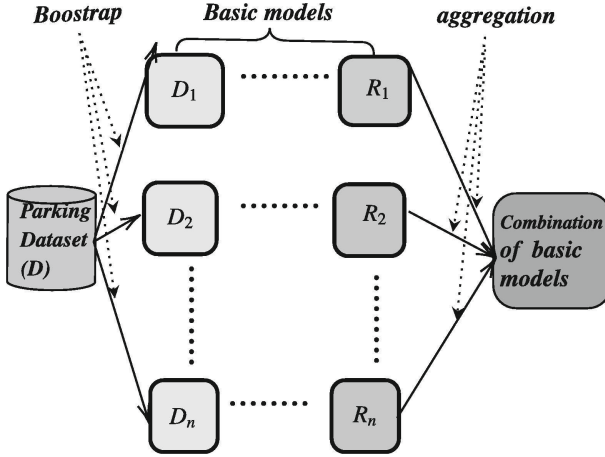


Fig. 5. Process of ensemble-based model

data D into n data D_1, D_2, \dots, D_n . From each data set D_i we will construct an intermediary regressor R_i and the final regressor will be an aggregation of the intermediate regressors R_i . From this general idea will be born several methods among which the most powerful are the bagging used in the Random forest algorithm and the Boosting used by Gradient Boosting and Adaptive Boosting.

4.3 Random Forest Regressor

Random forest is nothing more than a particular bagging method consisting of an aggregation of trees based on random variables. Most often, trees are built with the classification and regression tree (CART) algorithm whose principle is to recursively partition the space generated by the explanatory variables in a dyadic way. More precisely, at each stage of the partitioning, a part of the space is cut into two sub-parts according to a variable X_j .

Algorithm 1. Random Forest Regressor (RFR)

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

- 1: **for** $k = 1$ to B **do**
 - 2: Draw a bootstrap sample in d_n
 - 3: Construct a CART tree 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 $h(\cdot, k)$ the built tree.
 - 4: **end for**
-

Gradient Boosting Regression (GBR). Another very popular reinforcement algorithm is gradient enhancement. Gradient Boosting works similarly to AdaBoost by sequentially adding predictors to a set, so that everyone tries to correct the errors of its predecessor. However, instead of adjusting the instance weights at each iteration, as AdaBoost does, this method tries to fit the new predictor to the residual errors committed by the previous one [6].

5 Experimentation Results and Discussion

In this section, we will detail the experiments of the process of parking availability prediction based on ensemble method.

Algorithm 2. Gradient Boosting Regression (GBR)

Input: – x the observation to predict;

- h a weak rule,
- $d_n = (x_1, y_1), \dots, (x_n, y_n)$ the sample;
- λ a regularization parameter such as $0 < \lambda < 1$;
- M the number of iterations.;

Output: the estimator $\hat{g}_M(x) = \sum_{m=1}^M \alpha_m g_m(x)$

- 1: *Initialization* : Initialize the weight distribution of training data by $w_i = \frac{1}{N}, i = 1, 2, \dots, N$
- 2: **for** $m = 1$ to M **do**
- 3: Adjust the weak rule on the sample d_n weighted by the weights w_1, \dots, w_n , we note $g_m(x)$ the estimator resulting from this adjustment.
- 4: Compute $e_m = \frac{\sum_{i=1}^N w_i * 1_{y_i=g_m(x)}}{\sum_{i=1}^N n w_i}$
- 5: Compute $\alpha_m = \log\left(\frac{1 - e_m}{e_m}\right)$.
- 6: Readjust the weights:

$$w_i = w_i * \exp(w_i * 1_{y_i=g_m(x)}), i = 1, \dots, N$$

7: **end for**

5.1 Dataset

As shown in Fig. 4, our global predictive system consists of several phases. The first step consists to collect data from the sensors installed in the different smart-parkings. At this level, the data is collected in parking database as a csv file.

The data analyzed in this paper come from the Birmingham car park and were first used in (ref 1 and 2) comprising valid occupancy rates of 29 car parks operated by NCP (National Car Parks) in the city of Birmingham in the U.K. Birmingham, is a major city in the West Midlands of England, standing on the small River Rea. It is the largest and most populous British city outside London, with an estimated population of 1,124,569 as of 2016 [13]. Several cities in the

U.K. have been publishing their open data to be used, not only by researchers and companies, but also for citizens for better know the place where they live. The Birmingham data set is licensed under the Open Government License v3.0 and it is updated every 15 min from 8:00 AM to 4:30 PM (18 occupancy values per car park and day). In our study, we worked with data collected from Oct 4, 2016 to Dec 19, 2016 (11 weeks) which is available on UCI machine Learning Repository.

The selection of relevant data consists in eliminating irrelevant and redundant information. For the Birmingham Parking Database, the features considered relevant to the problem are: *SystemCodeNumber* : is an alphanumeric code that identifies a car park. *LastUpdated* : contains the date and time of the last update for occupancy data for each parking block. Schedules are recorded between 8:30 am and 6:30 pm. *Capacity* : Contains the capabilities of each car park. *Occupancy* : contains the occupations of each car park which are updated every 30 min. Other features such as the fill rate and the exit rate of each block were not considered in this work.

From these features we have generated a specific feature called the availability rate that we have noted *AVR* which is the ratio of the capacity minus the occupation at time t of the date d , on the capacity of the parking block. In our case, it is calculated by the following formula:

$$AVR_p(d, t) = \frac{Capacity_p - Occupancy_p(d, t)}{Capacity_p}$$

5.2 Performance Measures

In order to build an optimal approach, we compared the performance of different models using three main measures : the mean absolute error (MAE), the (R^2) and the root mean square error (RMSE). The three terms can judge the difference between the real and the predicted parking availability rate in different aspects. They are calculated as :

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (AVR_{ip} - AVR_i)^2}{N - 1}}$$

$$MAE = \frac{\sum_{i=1}^N |AVR_{ip} - AVR_i|}{N}$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (AVR_{ip} - AVR_i)}{\sum_{i=1}^N (AVR_{ip} - AVR_i)}$$

Where N is the total number of instances, AVR_{ip} is the predicted Availability Rate of the instance i and AVR_i is the real Availability rate of this instance.

The choice of a single measure may not always allow to separate the models. If the RMSE will show the error characterized by the variance and mean between

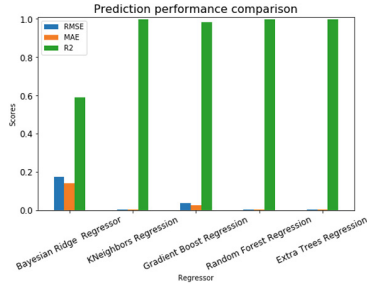


Fig. 6. Comparison of techniques

Table 1. Comparison of techniques

Regressor	RMSE	MAE	R2
Bayesian Ridge Regressor	0,1737	0,1416	0,5879
KNeighbors Regression	0,0014	0,0008	1,0000
Gradient Boost Regression	0,0348	0,0259	0,9835
Random Forest Regression	0,0017	0,0006	1,0000
Extra Trees Regression	0,0027	0,0010	0,9999

the predicted and the real by favoring the effects of the high deviations, the absolute error may reflect the effect of the precision in the prediction of the waiting time and R2 will show us the proportion of the actual waiting time that has been correctly predicted. The optimal model will result from the homogeneity between these three measures.

5.3 Results Analysis and Discussion

In order to find a model that gives the best prediction performance, we tested several regression models by starting with linear models (Bayesian Ridge Regressor) which did not prove very effective although very fast with RMSE of 0, 1737, MAE of 0.1416 and R2 of 0.5879. To improve these performances we have tested other stochastic and probabilistic models, so the results are shown in Fig. 6 and Table 1. According to these results KNeighbors and Random Forest give the best near-perfect performances reaching respectively 0.0014 and 0.0017 in terms of RMSE, 0.0008 and 0.0006 in terms of MAE and R2 of 100%. They are followed by Extratrees Regressor who gave RMSE performances of 0.0027, MAE of 0.0010 and R2 of 0.9999. With RMSE performance of 0.0348 MAE of 0.0259 and R2 of 0.9835, Gradient Boost Regression is in the last position just ahead of the much less efficient linear algorithms.

6 Conclusion

Urban mobility is one of the most important components interesting smart cities, and it's one of those It can directly benefit citizens. With the prediction of parking places, citizens they can reduce their time in the search for parking lot. In this paper we have proposed a prediction model of availability of parking spaces for the city of Birmingham. Specifically, the prediction techniques proposed showed better results than those of the baseline predictors.

References

1. Bélissent, J.: Getting clever about smart cities: New opportunities require new business models. Massachusetts, USA, Cambridge (2010)
2. Tang, S., Gao, H.: Traffic-incident detection algorithm based on nonparametric regression. *IEEE Trans. Intell. Transpo. Syst.* **6**, 38–42 (2005)
3. De Fabritiis, C., Ragona, R., Valenti, G.: Traffic estimation and prediction based on real time floating car data. In: 11th International IEEE Conference on Intelligent Transportation Systems, ITSC 2008 (2008)
4. Zheng, Y., Rajasegarar, S., Leckie, C.: Parking availability prediction for sensor-enabled car parks in smart cities. In: IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), p. 2015. IEEE (2015)
5. Tooraj, R., Ioannou, P.A.: On-street and off-street parking availability prediction using multivariate spatiotemporal models. *IEEE Trans. Intell. Transp. Syst.*, 2913–2924 (2015)
6. Lin, T., Rivano, H., Le Mouël, F.: A survey of smart parking solutions. *IEEE Trans. Intell. Transp. Syst.* **18**, 3229–3253 (2017)
7. De Almeida, P.R., et al.: PKLot-A robust dataset for parking lot classification. *Exp. Syst. Appl.* **42**, 4937–4949 (2015)
8. Lee, J.H., Hancock, M.G., Hu, M.C.: Towards an effective framework for building smart cities: lessons from Seoul and San Francisco. *Technol. Forecast. Soc. Change* **89**, 80–99 (2014)
9. Nam, T., Pardo, T.A.: Conceptualizing smart city with dimensions of technology, people, and institutions. In: Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times, pp. 282–291. ACM (2011)
10. Sharad, S., Sivakumar, P.B., Narayanan, V.A.: The smart bus for a smart city—a real-time implementation. In: 2016 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS), pp. 1–6. IEEE (2016)
11. Haitao, X., Ying, J.: Bus arrival time prediction with real-time and historic data. *Cluster Comput.* **20**(4), 3099–3106 (2017)
12. Rajabioun, T., Ioannou, P.A.: On-street and off-street parking availability prediction using multivariate spatiotemporal models. *IEEE Trans. Intell. Transp. Syst.* **16**(5), 2913–2924 (2015)
13. Camero, A., Toutouh, J., Stolfi, D.H., Alba, E.: Evolutionary deep learning for car park occupancy prediction in smart cities. In: International Conference on Learning and Intelligent Optimization, pp. 386–401. Springer, Cham, June 2018