EAI Endorsed Transactions
on Internet of Things
Research Article EAI.EU

Machine Learning Models in the large-scale prediction of parking space availability for sustainable cities

Abdoul Nasser Hamidou Soumana[1, *], Mohamed Ben Salah[2], Soufiane Idbraim[3], Abdellah Boulouz[4]

[1, 2, 3, 4]LabSiV Laboratory, Ibn Zohr University, Agadir, Morocco

Abstract

The search for effective solutions to address traffic congestion presents a significant challenge for large urban cities. Analysis of urban traffic congestion has revealed that more than 70% of it can be attributed to prolonged searches for parking spaces. Consequently, accurate prediction of parking space availability in advance can play a vital role in assisting drivers to find vacant parking spaces quickly. Such solutions hold the potential to reduce traffic congestion and mitigate its detrimental impacts on the environment, economy, and public health. Machine learning algorithms have emerged as promising approaches for predicting parking space availability. However, comparative studies on those machine learning models to evaluate the best suited for a large-scale prediction and within a given prediction time period are missing.

In this study, we compared nine machine learning algorithms to assess their efficiency in predicting long-term, large-scale parking space availability. Our comparison was based on two approaches: using on-street parking data alone and 2) incorporating data from external sources (such as weather data). We used automatic machine learning models to compare the performance of different algorithms according to the prediction efficiency and execution time. Our results indicated that the automated machine learning models implemented were well fitted to our data. Notably, the Extra Tree and Random Forest algorithms demonstrated the highest efficiency among the models tested. Moreover, we observed that the Random Forest algorithm exhibited less computational demand than the Extra Tree algorithm, making it particularly advantageous in terms of execution time. Therefore, this work suggests that the Random Forest algorithm is the most suitable machine learning model in terms of efficiency and execution time for accurately predicting large-scale, long-term parking space availability.

Keywords: Machine learning, Parking space prediction, Urban congestion reduction, Smart cities, Multi-output regression, Random Forest algorithm, Extra Tree algorithm

Received on 31 July 2023, accepted on 28 November 2023, published on 30 November 2023

doi: 10.4108/eetiot.2269

*Corresponding author. Email: abdoulnasser.hamidou@edu.uiz.ac.ma

1. Introduction

Recently, traffic congestion has become a major problem in large urban cities. Traffic congestion has proven to have a considerable impact on the various sectors of activity. In the United States, an economic loss of more than 101 billion dollars was due to traffic congestion in 2020. The losses were mainly attributed to increased fuel consumption and travel delays [1]. The combustion of fuel by vehicles leads to increased release of harmful substances proven to negatively impact health [2]. These substances have been linked to respiratory and cardiovascular diseases as well as premature deaths. According to the World Health Organization, China and India recorded more than 1 million and 600 thousand deaths related to traffic-related air pollution in 2012 [3]. Therefore, searching for solutions that will help reduce traffic-related air pollution is crucial.

Studies [4] have shown that more than 40% of traffic congestion in most large cities is due to the search for a parking space. Indeed, the inefficient search for available parking spaces contributes to longer search times and queues, further exacerbating traffic congestion. For example, in the United States, drivers spend an average of 17 hours per year searching for a parking space [5]. Several methods, such as machine learning, have been proposed for predicting the future occupation of parking spaces to effectively assist drivers in their search for parking spaces.

Our work provides a framework for comparing different machine learning models for predicting long-term parking space occupancy (i.e., 3 hours) in advance and on a large-scale. We used machine learning models for the long-term prediction of parking spaces through two approaches integrating on-street parking data from the city of Los Angeles and metrological data. Furthermore, a comparison of the different models' efficiency and computation time is also presented. Thus, we provided a reference for the most suitable machine learning algorithms for the long-term prediction that can serve as a basis for designing smart and sustainable parking solutions.

This article is organized as follows: the first section provides an overview of the current research on predicting parking space occupancy. Section 2 describes the dataset, machine learning algorithms, and methods used for modeling and prediction. In section 3, we present and discuss the results obtained by comparing the different methods in terms of their efficiency and computation time. Finally, section 4 concludes this work and gives future perspectives.

2. Related Work

Several methodologies have been suggested for predicting parking space availability. Many of these methodologies rely on or are compared to conventional machine learning models based on their performance. This prediction of parking space occupancy can be in the short-term (less than 30 minutes) or long-term (more than 30 minutes).

Sandeep Saharan et al. [6] conducted a comparative study of the performance of linear machine learning (LIN), decision tree (DT), neural network (NN), and random forest (RF) models for the implementation of intelligent on-street parking pricing system in the city of Seattle (Washington, USA). In a prediction of parking space one hour in advance, the Random Forest proved to be more efficient than the others in terms of precision. Goran Jelen and al. [7] evaluated the effectiveness of Random Forest and CatBoost machine learning models in predicting parking space occupancy one hour in advance using two approaches. In the basic approach, the model used parking space occupancy data only. On the other hand, the contextual approach utilized metrological data in addition to the data on parking space occupancy. The CatBoost proved to be more efficient than the random forest on all the approaches considered. In another study, Yanxu Zheng et al. [8] compared the performance of three machine learning models, regression tree, support vector regression, and neural network, using a set of short-term parking characteristics. Comparing the performance of these feature sets and model combinations for the Melbourne (Australia) and San Francisco (California, USA) datasets revealed that the regression tree with a feature set containing the historical observations, time of day, and day of the week provides the best performance.

However, these previous works have focused on prediction models based only on a limited number of parking lots and within a time frame of less than an hour. The analysis using automatic machine learning models on large-scale and long-term predictions of up to 3 hours has not been addressed. In addition, methodologies using machine learning methods as a benchmark have focused on a few prediction models, ignoring or limiting to less than 4 models benchmarking to find the models that perform the best within a given prediction time horizon.

In this work, we compare more than 10 machine learning models to determine the most efficient models, based on the accuracy and computation time in the prediction, that can be used as a reference for occupancy prediction parking spaces in the long term.

3. Motivation

3.1. Challenges of mobility and smart transportation

The transportation system is undoubtedly one of the main pillars of urban cities. The economy, viability, and development depend heavily on reliable transportation systems. In recent years, the evolution of urban cities toward smart cities has highlighted the need to address many transportation challenges to keep up with the dynamics and emergence of smart cities.

The integration of new information and communication technologies has led to the emergence of transportation systems toward more smarter systems (Figure 1).

Smart transportation systems are composed, on the one hand, of basic systems such as traffic control systems, circulation signs displaying various messages, and on the other hand, advanced systems combining several sources of information, including parking guidance information and parking reservation systems.

These systems are mainly based on innovative technologies such as:

- Sensor technologies (CCTV cameras, RFID sensors, magnetic sensors) interacting with their environments to collect real-time data on road conditions.
- Short- or long-range reach communication technologies, such as Wifi, 3G, 4G, and 5G, which allow data sharing.
- Wireless sensor networks are equipped with active or passive sensors communicating through low-power routing protocols to route the collected data.

The application fields of these technologies are numerous, including:

- ➢ Road safety
- ➢ Fluidification of urban traffic
- ➢ Resource management

- Environment protection

These challenges are intrinsically linked, particularly in the context of reducing traffic congestion in urban areas. This is mainly attributed to inadequate management of parking spaces, with adverse effects on traffic flow and urban infrastructure capacity.

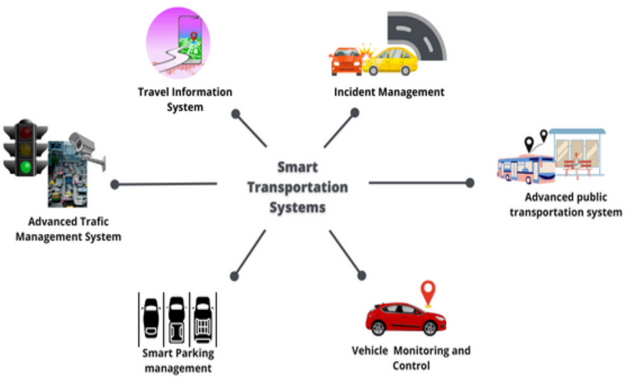

Figure 1. Intelligent transport system

3.2. Smart parking

Smart parking provides solutions to reduce traffic congestion and facilitate urban mobility. The smart parking system is composed of several components [9]. The most popular are illustrated in Figure 2. Among these components, the parking guidance and information system and the parking reservation system are vital in reducing urban traffic congestion.

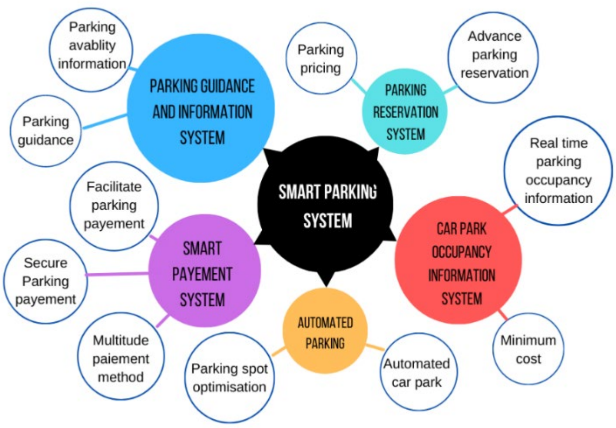

Figure 2. Smart parking system

The primary purpose of the parking guidance and information system (PGI) is to provide information on the availability of parking spaces and guidance in finding those available spaces. Such a system will allow drivers to spend less time looking for a parking space, reducing traffic congestion in certain areas. In recent years the information provided on the availability of parking spaces has become more and more accurate thanks to the collection of multi-source data and the use of deep learning techniques. For better applicability of the PGI system, several related questions need to be answered. Among these questions is how to avoid the race of several drivers toward the same parking space. The parking reservation system guarantees, in addition to intelligent pricing management, a booking management system including bookings for parking spaces, which helps avoid many vehicles continuing to circulate in search of available parking spaces. The intelligent management of parking spaces is made possible by using techniques to dynamically price parking spaces and make information on current availability more reliable. Most smart parking components mentioned previously, such as ERP and PRS, rely mainly on the efficient prediction of parking space availability. Making this prediction as effective as possible is inevitable for overcoming some of the significant transportation system challenges, such as traffic congestion in urban areas.

4. Materials and methods

We opted for two approaches to compare the various models considered in this study. The first approach evaluates the different algorithms based only on parking data. The objective of this first approach aims to analyze and identify the most efficient models based solely on historical parking data. The second approach explores how the first approach reacts to the integration of external data. In this second approach, our objective is to evaluate whether or not the result of the first approach remains valid even after integrating external data. The latter approach seems more critical because incorporating external data, such as meteorological data or traffic flow, to increase the efficiency of availability prediction has been suggested by some authors [7,10].

4.1. Data acquisition and pre-processing

The majority of methodologies suggested for predicting parking space availability have utilized historical data on parking space occupancy collected through various sources to construct the prediction models. However, studies have shown that the use of external data sources that can impact parking space occupancy, such as meteorological data, increases the prediction efficacy [7,10,11]. Therefore, in this study, we use historical on-street parking data from Los Angeles in the United States and incorporate the meteorological data as external data in our prediction models to examine how our models react to incorporating these data types.

Parking Data

Parking data were obtained from the Los Angeles Department of Transportation (LADOT) website [12]. LADOT manages all the on-street parking spaces throughout the city. We observed that on-street parking data from Los Angeles were recorded at irregular time intervals. Since we aim to make hourly predictions, we started by aggregating the parking data by hours. Then, each sensor represented by an ID is linked to a specific street using information about the geographical position of the sensors. Finally, at each street level and for each hour t, a value representing the number of free parking spaces (NPL) is calculated using the formula:

$$\text{NPLstreet(k)(t)} = \sum_{i+1}^{n} Espace_{(i)}(t) \qquad (1)$$

Where k represents the index of a specific street, n is the total number of parking spaces belonging to a street k, and Space(i), a value equal to either 0 or 1, which represents the state of a space of parking i at time t depending on whether the space is respectively occupied or free. This NPL quantity at each instant represents the value to be predicted for each hour on a given street.

Next, an aggregation process that consists in aggregating the arteries belonging to the same alley is carried out to reduce the relatively high number of streets. After aggregation, a total of 101 arteries is obtained, and the availability of parking spaces for these arteries will be the basis of our comparative study. Figure 3 shows an overview of data on the availability of parking spaces by hour on the first day of the year 2018 on three arteries in the city of Los Angeles

Meteorological data

Meteorological data for the city of Los Angeles were obtained using the Visual Crossing API [13]. These weather data were retrieved hourly and contained temperature, atmospheric pressure, and precipitation information. The only processing to be performed on the data was the conversion of the variables to the format specific to each variable. The weather information taken into account in our study were: temperature, precipitation, wind speed, atmospheric pressure, gusts of wind, and visibility.

Input data for the different models

The calculation of autocorrelation and partial autocorrelation based on the ARIMA methodology and inspired by previous work [10], showed that the information on the parking spaces availability of the last 24 hours is necessary to predict the next 3 successive hours. As illustrated in Figure 4, the parking data at the level of each artery is in the form of time series where each hour is linked to the availability of parking spaces. A sliding window of 24 elements with a time step of 1 traverses the time series to create the input matrix of the different models. For every 24 elements, extract the next 3 availabilities-hours representing the values to be predicted. For our first approach, at the level of each artery, for every 24 elements to extract defined as explanatory variables, the following 3 hourly occupations represent the target values to be predicted. For the second approach, the explanatory variables are defined by the 24 elements extracted plus the meteorological variables: temperature, precipitation, wind speed, atmospheric pressure, gusts of wind, and visibility. The set of all the explanatory variables extracted constitutes the input matrix of the models, and the set of all the following 3 hourly availabilities represents the matrix of values to be predicted.

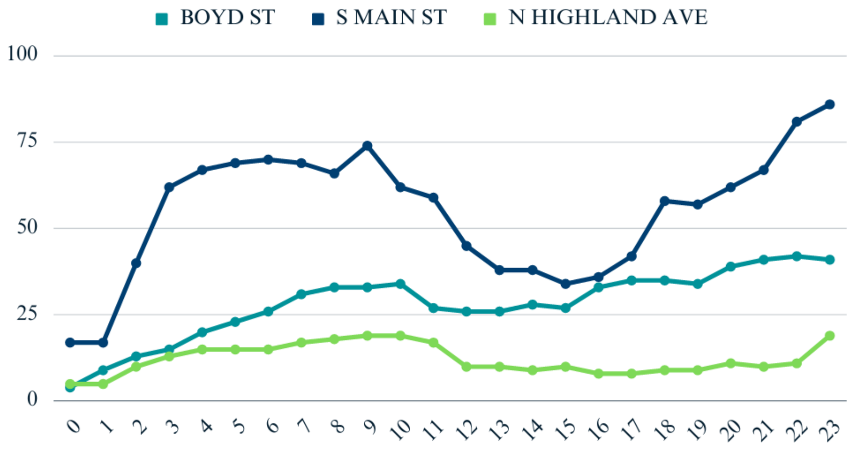

Figure 3. Overview of hourly parking availability at Boyd Street, South Main Street and North Highland Avenue.

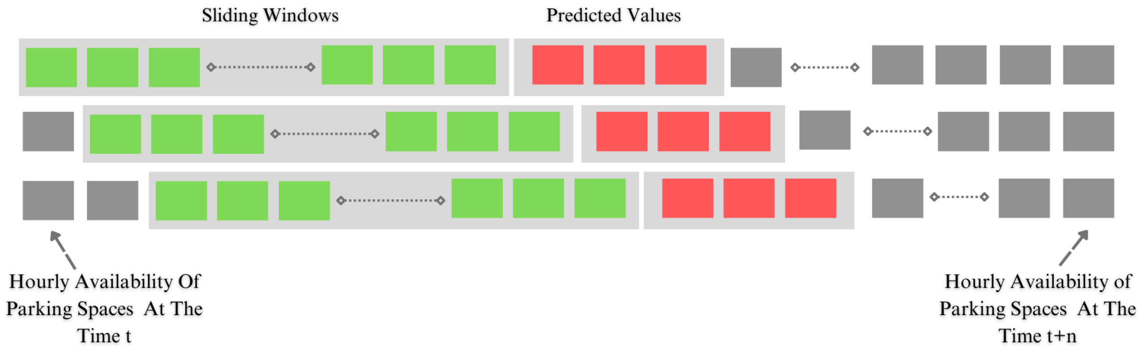

Figure 4. Illustration of parking availability time series data

Machine learning methods

The performances of various popular machine learning algorithms were compared based on their efficiency and training cost in prediction. In addition, the ability to predict occupancy for the next 3 hours was tested. Our choice is justified by the understanding that a long-term prediction, such as three hours in advance, can significantly benefit drivers in trip planning. By providing more advanced information, drivers gain valuable insights into the likelihood of finding an available parking space on a specific street, allowing them to plan their trips more effectively.

Multi-output prediction

The models compared in this study can be classified into two categories: the model intrinsically supporting multiple output prediction: Random Forest [14], K Neighbors Regressor [15], Linear Regressor [16], Ridge Regression [17], Decision Tree Regressor [18], and those that do not support this type of prediction: Extreme gradient boosting regression (XG-Boost) [19], Extra Tree [20], Gradient Boosting Regression[18], Stochastic gradient descent Regression[21], Support Vector Machine Regression (SVR) [22]. To overcome the limitation of these later models, we used an encapsulation technique, which consists of encapsulating several models to predict each element of the output sequence. In addition, encapsulating models can predict independent outputs or dependent outputs. For the models with independent outputs, each model independently predicts an element of the sequence according to the input data. In contrast, for the models with dependent output (Figure 5), the models receive the sequence element predicted by the previous model to predict the current sequence element during the prediction in addition to the input data.

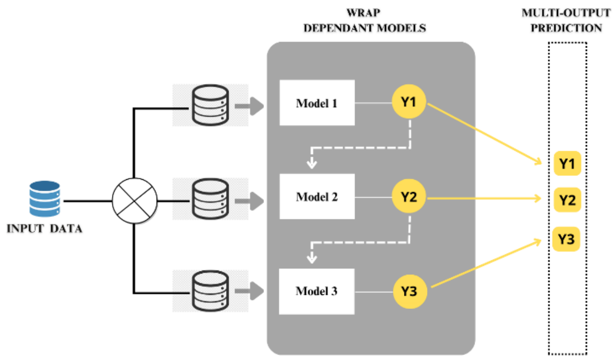

Figure 5. Encapsulation of models for multiple-output regression

Although in our case, the assumption that the sequence element is independent may not be correct because the parking availability for the next three hours probably has a dependency relationship between them. Nevertheless, we included both types of versions of each model in the comparison. The complete list of models used in this comparison is provided in Table 1.

Parameter optimization

Machine learning methods comprise numerous hyperparameters that require optimization to enhance the predictive model's efficiency. In this study, Bayesian optimization is employed to determine the optimal parameters. Initially, an exhaustive search is conducted for each model to identify the most significant parameters. Subsequently, Bayesian optimization is utilized to obtain the optimal values for the various parameters.

Table 1. Models and optimal parameters considered in the comparison

Model	Optimal Parameters
Ridge (RDG)	Default parameters
Decision Tree Regressor (DTR)	max_depth = 56, min_samples_split = 16, min_samples_leaf = 2, max_features = 17, max_leaf_nodes = 101
XG Boost with independent output (XGB1)	max_depth = 5, n_estimators = 100, gamma = 0.02
XG Boost with dependent output (XGB2)	
Extra Trees Regressor with independent output (EXT1)	Bootstrap = False, max_depth = 46, min_samples_leaf = 2, min_samples_split = 6, n_estimators = 651
Extra Trees Regressor with dependent output (EXT2)	
Gradient Boosting Regressor with independent output (GBR1)	random_state = 0, alpha = 0.3671, learning_rate = 0.08012, max_depth = 4, min_samples_split = 13, min_samples_leaf = 3, n_estimators = 271
Gradient Boosting Regressor with dependent output (GBR2)	
Stochastic gradient descent Regressor with independent output (SGDR1)	Alpha = 0.0, eta0 = 0.29, n_iter_no_change= 478, power_t = 0.3, tol = 0.58
Stochastic gradient descent Regressor with dependent output (SGDR2)	
SVM Regressor with independent output (SVR1)	C = 5.5, epsilon = 0.07, max_iter = 1570, tol = 0.1
SVM Regressor with dependent output (SVR2)	
Random Forest Regressor (RF)	max_depth = 10, min_samples_leaf = 2, min_samples_split = 4, n_estimators = 352

Table 1 presents a summary of the parameters utilized for each model, along with the values obtained through Bayesian optimization for each parameter. Certain models exhibited similar results irrespective of the considered parameter values. Hence, the default parameters were selected as the optimal parameters for these models.

5. Results and discussion

We have chosen the R2 and the RAE as evaluation metrics to compare the different models. The R2 metric shows how the model fits the prediction task. The formula for R2 is given by equation (2), a maximum R2 value of 1 indicates that the predictions of the regression model match the data perfectly. RAE indicates the error rate between the predicted and actual values, as defined by equation (3). Finally, k-fold cross-validation with k=5 is employed to assess the models and evaluate their robustness against data variations. This approach provides a reliable estimation of the models' generalization capability by considering different training and testing set configurations.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \quad (2)$$

$$RAE = \frac{\sum_{i=1}^{n}|yi - \hat{y}_i|}{\sum_{i=1}^{n}|yi - \bar{y}_i|} \quad (3)$$

Where ŷi is the predicted number of free parking spaces, yi is the actual number of available parking spaces, and $\bar{y}_i$ represents the average (mean).

We used the Python programming language and the sklearn library to implement the different learning models. For the encapsulation of the models, the MultiOutputRegressor and RegressorChain functions of the multioutput module of the sklearn library were used respectively for the models with independent outputs and the models with dependent outputs. The Google Colab environment was used as the runtime environment.
The average R2 and RAE provided by the different models during training on the parking data are presented in Table 2. We noticed on the parking data that all the models reached a minimum R2 of 0.6, proving that all the algorithms adapt well to the data. Among the models considered, the Extra Tree and the Random Forest were the ones that best adapted to parking space availability data with a maximum R2 of 0.70, i.e., 10% more than the minimum R2 reached by the Ridge. By comparing the RAE of the different algorithms, we also noticed that the latter algorithms provided the best accuracy with the lowest error rate. On the other hand, we observed that the encapsulated models with dependent and independent output gave approximately the same result. This proved that although the different models predicted the future availability of three consecutive hours probably dependent on each other, taking this information into account did not significantly improve the prediction.

Table 2. Model's performance

Model	Parking		Parking + weather	
	Mean R2	Mean RAE	Mean R2	Mean RAE
RDG	0.66	0.49	0.63	0.50
DTR	0.60	0.51	0.59	0.52
XGB2	0.69	0.45	0.68	0.46
XGB1	0.69	0.45	0.68	0.46
EXT1	**0.70**	**0.44**	**0.70**	**0.45**
EXT2	**0.70**	**0.44**	**0.70**	**0.45**
GBR1	0.68	0.46	0.67	0.47
GBR2	0.68	0.45	0.67	0.47
SGDR1	0.66	0.49	0.47	
SGDR2	0.66	0.49	0.47	
SVR1	0.65	0.49	0.61	0.50
SVR2	0.65	0.50	0.61	0.51
RF	**0.70**	0.45	**0.70**	**0.45**

Observing the R2 and the average RAE of the different algorithms following the incorporation of meteorological data (Table 2), the Extra Tree and the Random Forest are also the models that fitted the data the best and provide the best prediction even after the incorporation of external data.

It is noted that the two versions of SVR did not provide results after the incorporation of meteorological data. Therefore, these results are not presented in Table 2.

In addition, the comparison of the results of the models with meteorological data and those without meteorological data (Table 2) shows a decrease in R2 and RAE after incorporating meteorological data for particular models. This proves that the incorporation of meteorological data does not contribute to the improvement of the prediction. By comparing the execution time of the different algorithms, we found that the Extra Tree (Figure 6) is the algorithm that takes a higher execution time due, in particular, to the relatively high number of parameters to train during the training phase. Therefore, the Random Forest algorithm is more adopted than the Extra Tree in efficiency and computational cost.

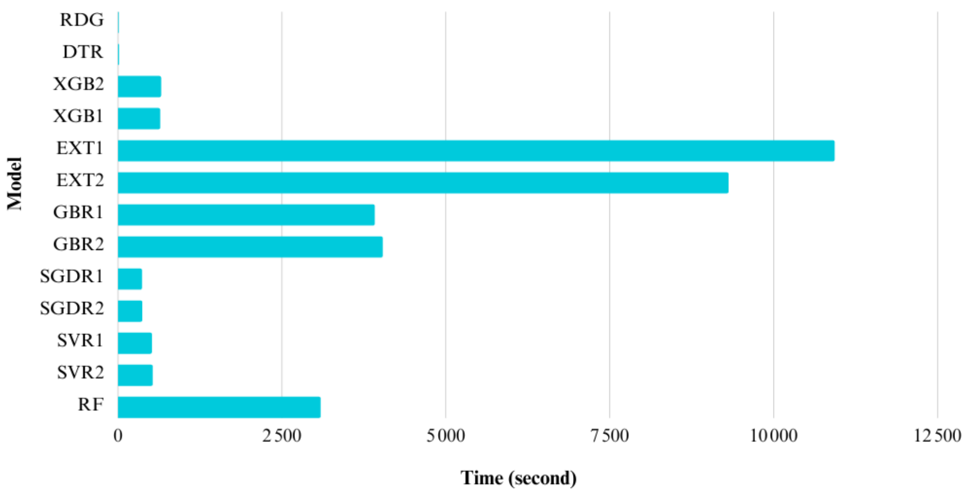

Figure 6. Execution cost of each model per second

6. Conclusion and perspective

Major urban cities face several challenges related to traffic congestion. Part of this congestion in urban areas is attributed to the unassisted and prolonged search for parking spaces, which frequently leads to frequent traffic jams. To solve this problem, predicting parking space availability appears to be a promising alternative to make parking management smarter and facilitate smoother urban road traffic. In this regard, several solutions have been developed to address the congestion problem associated with parking. These solutions involve predicting current and future availability and providing users with information to find vacant parking spaces easily. Adapting such approaches will reduce search time for parking spaces while improving traffic flow, minimizing fuel consumption, and reducing traffic-related air pollution in urban areas. In this study, we compared the performance of machine learning algorithms for large-scale and long-term parking availability prediction. Based on two approaches incorporating multi-output prediction techniques and data from the city of Los Angeles, our comparative study evaluated the performance of over 10 machine learning algorithms. Our experiments revealed that Extra Trees and Random Forest algorithms outperformed other models in predicting parking space availability. Furthermore, comparing the different algorithms in terms of execution time showed that Random Forest had the best execution time compared to Extra Trees. Several authors have highlighted the possibility of integrating external data sources for improved prediction accuracy. Nevertheless, our experiments revealed that incorporating weather data did not enhance prediction. In future work, we plan to explore integrating other external data sources, such as event data, to assess their effect on predicting parking availability. Also, with the advancements in deep learning and its tremendous success across various domains, such as the prediction of time series, we envision the emergence of algorithms based on this type of learning for large-scale parking availability prediction. Thus, it will be interesting to compare the performance of deep learning models with that of Random Forest and Extra Trees algorithms.

Acknowledgements.
The authors would like to thank the Los Angeles Department of Transportation (LADOT) for providing free access to on-street parking data for the City of Los Angeles.

References

[1] D. Schrank, B. Eisele, T. Lomax, 2021 Urban Mobility Report, (2021).
[2] R. Doolan, G.-M. Muntean, VANET-Enabled Eco-Friendly Road Characteristics-Aware Routing for Vehicular Traffic, in: 2013 IEEE 77th Vehicular Technology Conference (VTC Spring), 2013: pp. 1–5.
[3] S.S. Pulugurtha, V.R. Duddu, M. Venigalla, Evaluating spatial and temporal effects of planned special events on travel time performance measures, Transportation Research Interdisciplinary Perspectives. 6 (2020) 100168.
[4] S. Kazi, S. Nuzhat, A. Nashrah, Q. Rameeza, Smart Parking System to Reduce Traffic Congestion, in: 2018 International Conference on Smart City and Emerging Technology (ICSCET), 2018: pp. 1–4.
[5] INRIX Economic Cost of Parking Pain Report, https://www2.inrix.com/research-parking-2017.
[6] An efficient smart parking pricing system for smart city environment: A machine-learning based approach, uture Generation Computer Systems, 2020: pp. 622-640.
[7] G. Jelen, V. Podobnik, J. Babic, Contextual prediction of parking spot availability: A step towards sustainable parking, Journal of Cleaner Production. 312 (2021) 127684.
[8] Y. Zheng, S. Rajasegarar, C. Leckie, Parking availability prediction for sensor-enabled car parks in smart cities, in: 2015 IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 2015: pp. 1–6.
[9] F. Al-Turjman, A. Malekloo, Smart parking in IoT-enabled cities: A survey, Sustainable Cities and Society. 49 (2019) 101608. https://doi.org/10.1016/j.scs.2019.101608.
[10] J. Arjona, M.P. Linares, J. Casanovas, A deep learning approach to real-time parking availability prediction for smart cities, in: Proceedings of the Second International Conference on Data Science, E-Learning and Information Systems, Association for Computing Machinery, New York, NY, USA, 2019: pp. 1–7.
[11] S. Yang, W. Ma, X. Pi, S. Qian, A deep learning approach to real-time parking occupancy prediction in spatio-temporal networks incorporating multiple spatio-temporal data sources, ArXiv:1901.06758 [Cs, Stat]. (2019). http://arxiv.org/abs/1901.06758 (accessed October 13, 2021).
[12] Los Angeles - Open Data Portal, City of Los Angeles. (n.d.). https://data.lacity.org/browse?category=transportation (accessed April 17, 2022).
[13] Weather Data Services | Visual Crossing, (n.d.). https://www.visualcrossing.com/weather/weather-data-services#/login (accessed October 13, 2021).
[14] L. Breiman, Random Forests, Machine Learning. 45 (2001) 5–32.
[15] F. Martínez, M.P. Frías, M.D. Pérez, A.J. Rivera, A methodology for applying k-nearest neighbor to time series forecasting, Artif Intell Rev. 52 (2019) 2019–2037.
[16] D. Maulud, A.M. Abdulazeez, A Review on Linear Regression Comprehensive in Machine Learning, Journal of Applied Science and Technology Trends. 1 (2020) 140–147.
[17] E. Cule, M. De Iorio, Ridge Regression in Prediction Problems: Automatic Choice of the Ridge Parameter, Genetic Epidemiology. 37 (2013) 704–714.
[18] J. Singh Kushwah, A. Kumar, S. Patel, R. Soni, A. Gawande, S. Gupta, Comparative study of regressor and classifier with decision tree using modern tools, Materials Today: Proceedings. (2021).
[19] J.H. Friedman, Greedy Function Approximation: A Gradient Boosting Machine, The Annals of Statistics. 29 (2001) 1189–1232.
[20] P. Geurts, D. Ernst, L. Wehenkel, Extremely randomized trees, Mach Learn. 63 (2006) 3–42.
[21] A. Sharma, Guided Stochastic Gradient Descent Algorithm for inconsistent datasets, Applied Soft Computing. 73 (2018) 1068–1080.
[22] M. Sabzekar, S.M.H. Hasheminejad, Robust regression using support vector regressions, Chaos, Solitons & Fractals. 144 (2021) 110738.