Short-term prediction of parking availability in an open parking lot

Vijay Paidi[1]

[1] Department of Computer and Information Management, School of Information and Engineering, Dalarna University Borlänge, Sweden

Abstract

The parking of cars is a globally recognized problem, especially at locations where there is a high demand for empty parking spaces. Drivers tend to cruise additional distances while searching for empty parking spaces during peak hours leading to problems such as pollution, congestion, and driver frustration. Providing short-term predictions of parking availability would facilitate the driver in making informed decisions and planning their arrival to be able to choose parking locations with higher availability. Therefore, the aim of this study is to provide short-term predictions of available parking spaces with a low volume of data. The open parking lot provides parking spaces free of charge and one such parking lot, located beside a shopping center, was selected for this study. Parking availability data for 21 days was collected where 19 days were used for training, while multiple periods of the remaining 2 days were used to test and evaluate the prediction methods. The test dataset consists of data from a weekday and a weekend. Based on the reviewed literature, three prediction methods suitable for short-term prediction were selected, namely, Long-short term memory (LSTM), Seasonal autoregressive integrated moving average with exogenous variables (SARIMAX), and the Ensemble-based method. TheLSTM method is a deep learning-based method, while SARIMAX is a regression-based method, and the Ensemble method is based on decision trees and random forest to provide predictions. The performance of the three prediction methods with low volume of data and the use of visitor trends data as an exogenous variable was evaluated. Based on the test prediction results, the LSTM and Ensemble-based methods provided better short-term predictions at multiple times on a weekday, while the Ensemble-based method provided better predictions over the weekend. However, the use of visitor trend data did not facilitate improving the predictions of SARIMAX and the Ensemble-based method, while it improved the LSTM prediction for the weekend.

1. Introduction

Parking of personal vehicles around areas of public interest is a well-known global problem, and it exacerbates when there is a smaller number of available empty parking spaces for many vehicles. The availability of empty parking spaces varies over space and time. This problem is often encountered at locations such as airports, shopping centers, etc., where there is high demand for empty parking spaces (Tiedemann et al., 2015). Lack of a sufficient number of empty parking spaces during peak hours leads to problems such as congestion, pollution, excess cruising, and driver frustration (Paidi et al., 2022, Song et al., 2019). Approximately, 30% of congestion in urban areas was caused by drivers looking for empty parking spaces (Shoup, 2018). Population in urban areas is expected to increase by 12% until 2050, which exacerbates these parking associated problems.

Excess cruising and other parking-associated problems occur primarily due to lack of information on parking availability. According to (Van Der Waerden et al., 2011), parking spaces are better utilized if occupancy information is available to the driver. Real-time parking availability can be displayed to the driver either using parking guidance systems or smart parking applications. However, such systems are only available for closed parking lots where there is a return on investments. Closed parking lots provide paid parking spaces, while open parking lots provide free parking spaces for a limited duration of time. There are no similar applications available for open parking lots due to a lack of return on investments (Paidi et al., 2018). Providing real-time occupancy information for open parking lots would not be of much use as parking lots are available only for a short duration due to high demand (Rajabioun and Ioannou, 2015). However, providing short-term predictions of parking spaces would facilitate the driver in making informed decisions on their arrival at the parking lot (Caicedo et al., 2012). Thus, this study aims to provide a short-term prediction of parking availability in an open parking lot using limited historical data. Disseminating short-term predictions of parking availability would help the driver to avoid peak times and choose parking areas with higher availability and reduce additional cruising. Reduction in cruising for empty parking spaces also reduces CO_2 and other harmful emissions (Caicedo, 2010). There are not many previous studies providing short-term predictions of parking availability for open parking lots primarily due to lack of return in investments. This paper facilitates in addressing this research gap.

Historical parking availability data needs to be collected to train forecasting methods. It is not economically viable to install underground sensors at open parking lots since sensors incur high maintenance costs (Paidi et al., 2018). An optical camera facilitates in capturing collective parking occupancy data. However, due to issues with lighting conditions and privacy, a thermal camera was utilized to collect data at an open parking lot. Generally, historical data spanning over months or years would be used to train forecasting methods. However, to evaluate short-term predictions with a low volume of data, parking availability data, collected over 21 days was utilized in this study. The parking availability data was collected using a thermal camera at an open parking lot beside a shopping center.

Previous studies like (Zhao and Zhang, 2020), focused on predicting parking spaces at commercial spaces or office spaces. Similarly studies like (Rajabioun and Ioannou, 2015, Pozo et al., 2021), evaluated prediction methods using large volume of parking data collected from parking providers. However, there were not many studies predicting parking spaces at open parking lots. Parking availability differs from each location of interest, such as restaurants, shopping centers, residential areas, or office buildings. There can be several variables which facilitate in understanding the parking availability, such as holidays, super sales, day of the week, office hours, etc. Therefore, previous studies utilize some of the relevant and available variables to improve the prediction performance of algorithms. In a study by (Richter et al., 2014), historic trends data was utilized to improve the accuracy of the forecasting model. However, to capture historical trends data, a large volume of data is necessary. Therefore, in this study, since there was a limited volume of data available, historical trends were categorized using visitor trends data from Google popular times. Regression-based machine learning algorithms are suitable for stationary data (Shao et al., 2018). The collected parking availability data has a constant mean and variance over a period of time and is considered stationary. Therefore, a regression-based algorithm, namely Ensemble, was utilized in this study. Recurrent neural network-based algorithms, such as LSTM, are deep learning-based forecasting methods suitable for short-term predictions, capable of learning complex patterns, and therefore utilized in this study to predict parking availability. LSTM is suitable for forecasting stationary and non-stationary data. Similarly, SARIMAX is another suitable method for short-term predictions. It learns patterns from previous values considering seasonality and exogenous variables (Fokker et al., 2021). In this study, exogenous variables, such as time, day of the week, visitors trend from Google, holidays, and rain or snow, were utilized to facilitate forecasting. These are further discussed in the Method section.

The contribution of this paper is as follows.

- The paper provides short-term predictions using a limited volume of parking availability data collected on an open parking lot. T
- This paper evaluates the use of visitor trend data from Google, which is an open-source visitor trends data.
- Finally, prediction methods such as LSTM, SARIMAX, and an Ensemble-based method were evaluated to provide short-term predictions at multiple times of the day.

The remaining sections of the paper are organized as follows. Section 2 discusses relevant literature on predicting parking availability. Section 3 focuses on the data collection and prediction algorithms. Section 4 presents results and discussion, while the paper ends with a conclusion in Section 5.

2. Related work

The prediction of parking spaces in previous literature was performed using several methods, such as neural networks, multivariate spatiotemporal models, deep learning, and machine learning algorithms. This section discusses these methods and their suitability for this paper.

In (Rajabioun and Ioannou, 2015), parking lot forecasting of on-street and off-street parking was performed using a multivariate model. The on-street and off-street parking lots are paid parking spaces that fall under the category of a closed parking lot. The data for the study was obtained from SFpark. The multi-variate model also utilized time variants, such as accidents and maintenance, in forecasting parking occupancy information. The model achieved a 95% accuracy for a 20-minute forecast horizon. It performed well with off-street parking lots compared to on-street parking lots, which have more variance. The model needs several spatial and temporal variables to forecast parking occupancy information. Also, the model is suited for large cities with several parking lots within the vicinity. The data provided by SFPark consists of additional variables which were not available for the open parking lot dataset collected in this paper.

In (Rajabioun et al., 2013), a forecast of free occupancies was provided based on drivers preferences, such as cost, walking distance, parking rules, duration, and availability. Preferred parking location was forecast based on cost function. A forecast of parking occupancy was provided based on historical and real-time data. The cost is not available for open parking lots as parking spaces are provided free of charge. Therefore, the model proposed in (Rajabioun et al., 2013), is not suitable for this study.

Information on historical parking availability with 5-minute interval over 5 months was utilized in (Richter et al., 2014). The clustering method was used to reduce the storage of data in providing forecasting information. Due to this, an accuracy level of 70% was achieved in its predictions. The average values along with categories of availability were utilized for training the algorithm. Categories of availability were low, medium, and high. Utilizing average values facilitates improving accuracy, even during holidays. Therefore, this paper also employs a similar approach, utilizing average occupancy values and availability categories.

Simple recurrent neural networks have problems with exploding and vanishing gradients, unlike LSTM's. The architecture of LSTM consists of a memory cell that overcomes the vanishing gradient problem and facilitates learning sequential data (Hochreiter and Schmidhuber, 1997, Fan et al., 2020). In (Shao et al., 2018), a prediction of parking availability was performed using LSTM and clustering techniques. Parking occupancy and duration were provided as input for the algorithm. A large parking lot was divided into clusters with similar patterns and recurrent neural network is used to generate forecasting metrics which is fed into the parking model. The proposed model is not suitable for this study as only one region of a parking lot was selected, and duration of the parking was not compiled. In (Fan et al., 2020), a multi-step prediction approach using LSTM was utilized to predict long-term parking availability information. Large amounts of historical data were utilized to predict parking availability for 60 minutes. The predictions of LSTM performed better compared to regression-based models. Similarly, LSTM was utilized in this study as well, to provide predictions for 60 minutes but with data available for three weeks.

As discussed in (Fokker et al., 2021), the parking availability forecast for 6 months ahead was performed using SARIMAX and LSTM algorithms. SARIMAX performed better than LSTM, leading to lower RMSE values. The inclusion of exogenous variables resulted in reducing errors up to 27%. Hence, SARIMAX was also utilized in this study to perform forecasting of parking availability. In another study,Fışkın and Cerit (2019), SARIMAX was utilized to predict cement demand. The data consists of 178 points which were divided to train and test datasets. SARIMAX produced fewer errors in predicting cement demand. In (Fokker et al., 2021, Fışkın and Cerit, 2019) studies, large volume of data was utilized to train SARIMAX method. However, in this paper, limited volume of data was utilized which might impact the accuracy of predictions.

The Ensemble-based model is another popular method used for prediction and classification. In (Tekouabou et al., 2020), the ensemble method was used to predict the availability of empty parking spaces, achieving a Mean Absolute Error (MAE) value of 0.06 in its prediction. Bagging and boosting approaches were evaluated in this study. Similarly, in another study (Awan et al., 2020), an ensemble approach combines multilayer perceptron, K-nearest neighbors, decision trees, and random forest to predict parking occupancy. A voting classifier was utilized to make decisions based on individual predictions. Less complex learners, such as decision trees, random forest, and KNN, performed better, compared to multilayer perceptron. Hence, the ensemble-based method with bagging approach is also utilized in this study to predict parking availability.

There were no studies that collected data from an open parking lot. Several previous studies utilized data collected from closed parking lots where sensors and other tools were used to obtain parking availability and duration information for larger periods, such as months or years. However, in this study, the data was collected manually using a thermal camera and thus the volume of data is limited. The relevant variables utilized in this paper are open-source data and not collected using external sensors or sources. Popular prediction algorithms such as LSTM, SARIMAX, and Ensemble methods were suitable for prediction purposes and, thus, they were utilized and evaluated in this study.

3. Materials and Method

This section discusses the location of the parking lot and the method utilized to predict the availability of parking occupancy.

3.1 Cast study

The parking lot is located beside a shopping center in a mid-sized city of Sweden. A fraction of that parking lot is the region of interest (ROI) in this study, and it is highlighted by a green color line, as illustrated in Figure 1. A thermal camera is installed in the shopping center and is utilized to collect videos from the ROI. It was possible to collect data from an open parking lot without heavy installation and maintenance costs due to the use of the thermal camera. The thermal camera also avoided privacy concerns that would have been caused by the use of an optical camera (Asghar et al., 2019). The data for the entire parking lot could not be considered due to the limited field view of the camera. The ROI has four entrances, E1, E2, E3, and E4, and since it is located near the entrance of the shopping center, it is expected to have a higher vehicle traffic (Paidi, 2021). The thermal camera views the vehicle from the side instead of the rear or front side of the vehicle, which hinders the visibility of the vehicle. Therefore, the parking occupancy data was collected manually.

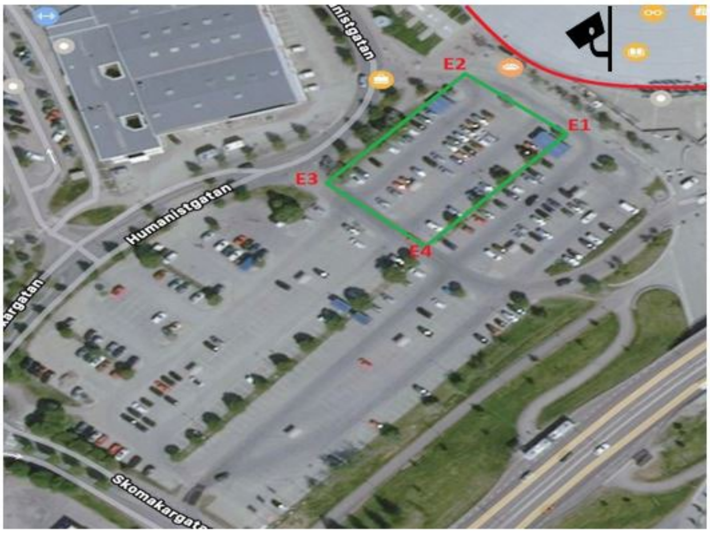


Figure 1. Observed parking lot and the region of Interest highlighted by green line

3.2 Data collection:

There are previous studies that forecast parking occupancy information using months of historical data combined with real-time parking occupancy data. Such data can be collected for closed parking lots, where underground sensors can be utilized to collect long-term data. However, it is expensive to obtain such data for an open parking lot due to lack of Parking Guidance and Information (PGI) systems. Therefore, data collection was performed using videos from a thermal camera installed on the dome of the shopping center. Parking occupancy data was extracted manually using these videos. The number of parking spaces covered in this study is 47. The data for the forecast is collected over 21 days, between January 7 and January 27, 2020, as shown in Table 1. The first 19 days of data were used to train the model, while the remaining two days of data were used to evaluate the model. As mentioned in (Balmer et al., 2021), at least 5-days of hourly data is necessary to provide reasonable prediction values. Hence, the allocated training data size was sufficient to produce better prediction values. The parking spaces in an open parking lot are in higher demand and frequent changes within short intervals are expected. Therefore, the parking occupancy data were collected every 5-minutes. Since this paper utilizes limited data for parking availability forecasting, it uses additional exogenous variables to improve the accuracy of forecasting. Exogenous variables as shown in Table 1 were utilized in this paper. This consists of a total of 6 columns where the parking availability output variable is to be forecasted and the remaining are exogenous variables or predictors. Variables, such as level of traffic to the shopping center, day, holiday, and rain or snow, are exogenous as they are not dependent on the availability of parking spaces.

Table 1. Metadata of the dataset

Date	**Day**	**Holiday**	**Date**	**Day**	**Holiday**
Jan-07-2020	Tuesday	No	Jan-18-2020	Saturday	No
Jan-08-2020	Wednesday	No	Jan-19-2020	Sunday	No
Jan-09-2020	Thursday	No	Jan-20-2020	Monday	No
Jan-10-2020	Friday	Yes	Jan-21-2020	Tuesday	No
Jan-11-2020	Saturday	No	Jan-22-2020	Wednesday	No

Jan-12-2020	Sunday	Yes	Jan-23-2020	Thursday	No
Jan-13-2020	Monday	Yes	Jan-24-2020	Friday	No
Jan-14-2020	Tuesday	No	Jan-25-2020	Saturday	No
Jan-15-2020	Wednesday	No	Jan-26-2020	Sunday	No
Jan-16-2020	Thursday	No	Jan-27-2020	Monday	No
Jan-17-2020	Friday	No			

Table 2. Example of time series input data for the model

Local Time	Availability (%)	Level	Day	Holiday	Rain or Snow
2020-01-07 09:00	77	Low	Tuesday	No	No
2020-01-07 9:05	75	Low	Tuesday	No	No
2020-01-07 9:10	72	Low	Tuesday	No	No

- Local time: This column provides the timestamp of the collected data, where the date and time are displayed. The data, as shown in Table 2, is captured between 9:00 and 20:00, with a frequency of 5 minutes. Each day represents 133 data points.
- Availability: This column is the parking spot availability represented in percentage, i.e. $Availability = \frac{max\{E_j\}}{47} \times 100; j \in w_i$, where E_j is the number of empty parking spots out of 47 available spots at any point in time j, when the change in the empty parking spots occurs within a specific 5 minutes, the observation window w_i, i=1, 2, n. During peak periods, the availability becomes low, and during non-peak periods, it would be high. This is also the output or forecasted variable. The data on availability rate is illustrated through a time series plot in Figure 2.

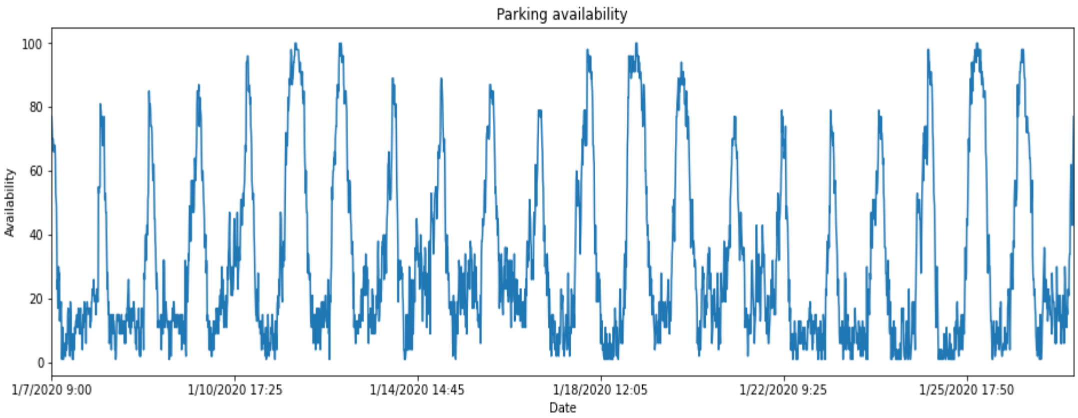


Figure 2: Illustrates parking availability of the total dataset

- Level: This column is based on the visitor trend data from Google. However, the collected data is restricted to people using Google-based applications, such as Google Maps. Parking occupancy differs from each location of interest, such as restaurants, shopping centers, residential areas, or office buildings. Therefore, as mentioned in (Ionita et al., 2018), usage of parking profiles would facilitate improving the accuracy of the forecasting model. Popular times data is not available for certain periods, such as between 9:00 and 10:00, or between 19:00 and 20:00, due to few people visiting the shopping center. Since this only represents people using Google-based applications, these intervals are labeled as Low. The periods where visitors were less than 60% are designated as Medium. The periods near and around the peak, between 60% and 100%, are labeled as High. This variable is exogenous as visitor traffic may, or may not, affect parking availability. While visitor traffic is similar on weekdays, it varies on the weekend, as shown in Table 3.

Table 3. Hourly level data during weekday and weekend

Hour	Friday	Saturday
9	Low	Low
10	Medium	Medium
11	High	Medium
12	High	High
13	High	High
14	High	High
15	High	High
16	High	High
17	High	Medium
18	Medium	Low
19	Low	Low
20	Low	Low

- Day: This column represents days of the week. The use of this column facilitates forecasting parking availability based on the day of the week, and not on timestamp. As shown in Figure 3, the availability decreases by 12:00 during weekdays, while it increases by 13:00 during weekends. Availability during weekdays also sharply increases after 17:00, but gradually increases from 16:00 over the weekend.

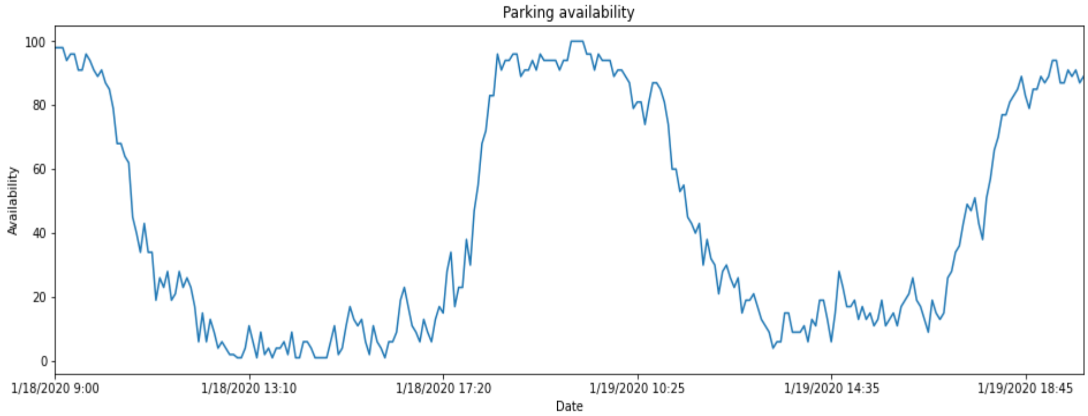


Figure 3. Parking availability during a weekday (1/18/2020) and a weekend (1/19/2020)

- Holiday: This column represents public holidays in Sweden. Parking occupancy varies on holidays. Therefore, it is necessary to consider this variable to forecast parking availability during holidays. However, traffic variation during all the holidays is not the same. Traffic during Christmas is high when compared to other public holidays. This variation is not observed in this study, due to the limited dataset.
- Rain or snow: This column describes if there is rain or snow. Weather conditions usually impact travel behavior. However, in a country like Sweden, we expect this impact to be minimal unless there is rain or snowstorm.

Approximately 2400 data points were utilized to train the selected algorithms, while 400 data points were utilized to evaluate the predictive performance of the algorithms. Training and testing were performed with and without visitor trend data to evaluate its effect on the performance of the algorithms. Two sets of prediction methods were trained, where one set includes visitor trend data during training and testing, while the other set does not include visitor trend data. Each set consists of LSTM, SARIMAX and Ensemble-based methods. The exogenous variables, such as holiday, rain or snow, and day of the week, are included as they were found to improve the accuracy of the algorithms in several previous studies (Basu and Little, 2004, Ghosal et al., 2019, Provoost et al., 2020, Fabusuyi et al., 2014). As parking availability varies between weekdays and weekends, the test dataset consists of data from a weekday and weekend. Prediction methods were evaluated at multiple times, such as 09:00, 13:00, and 18:00 during these days. Dell workstation with Quardro P5200 GPU was utilized in this study along with Matlab and Anaconda platforms to prepare data and evaluate prediction methods.

3.3 Algorithms for short-term prediction

Multiple algorithms, such as LSTM, SARIMAX, and Ensemble-based methods, are evaluated in this paper with and without visitor trend data. The exogenous variables mentioned in Section 3.1 are the predictors, while the parking availability information is the response of the prediction methods.

Long-short term memory (LSTM): Recurrent neural networks are well known for prediction capability, as the algorithm understands the patterns in sequential data. LSTM is a recurrent-based neural network that consists of a memory cell and is ideal for prediction purposes (Shao et al., 2018). This architecture consists of sequential, LSTM, dropout, and fully connected layers. Sequential data was fed into multiple LSTM and dropout layers. LSTM layer was used to learn patterns in sequential data, and the dropout layer was used to avoid overfitting (Fan et al., 2020). Finally, the fully connected layers were utilized to produce the output. The batch size of the LSTM is 12, while the optimizer is stochastic gradient descent. Since the data has a limited volume, the number of epochs for training is 10.

Seasonal autoregressive integrated moving average with exogenous variables (SARIMAX): SARIMAX is a time series model which incorporates seasonality and exogenous variables in forecasting. Like ARIMA, it consists of an autoregressive polynomial of order p, moving average polynomial of order q, and non-seasonal difference d to produce stationarity. Due to the inclusion of seasonality, it consists of a seasonal autoregressive polynomial of order P, a seasonal moving average polynomial of order Q, and seasonal difference D, while the length of periodicity is given by m (Fışkın and Cerit, 2019, Cools et al., 2009). Seasonal and non-seasonal difference is zero as the data is stationary. In this study, the length of periodicity is 12, as each hour consists of 12 data points.

Ensemble-based model: This Ensemble method utilizes the bagging approach which combines several decision trees, and random forests to individually provide predictions (Tekouabou et al., 2020). A weak learner, such as a decision tree, is taken over multiple selected samples and learns from them independently. Finally, it combines all the predictions based on deterministic average, and weights, to predict output. The minimum leaf size is between 1 and 1197, the number of learners is from 10-500, and the learning rate is 0.001.

3.3 Evaluation metrics

Parking availability predictions were provided for 60 minutes. The predicted results were evaluated using the following metrics, where y_i is the actual value at *i* the instance, while $\hat{y}_i$ is the predicted value at instance *i.*

Mean Absolute Error (MAE):

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

Root Mean Squared Error (RMSE):

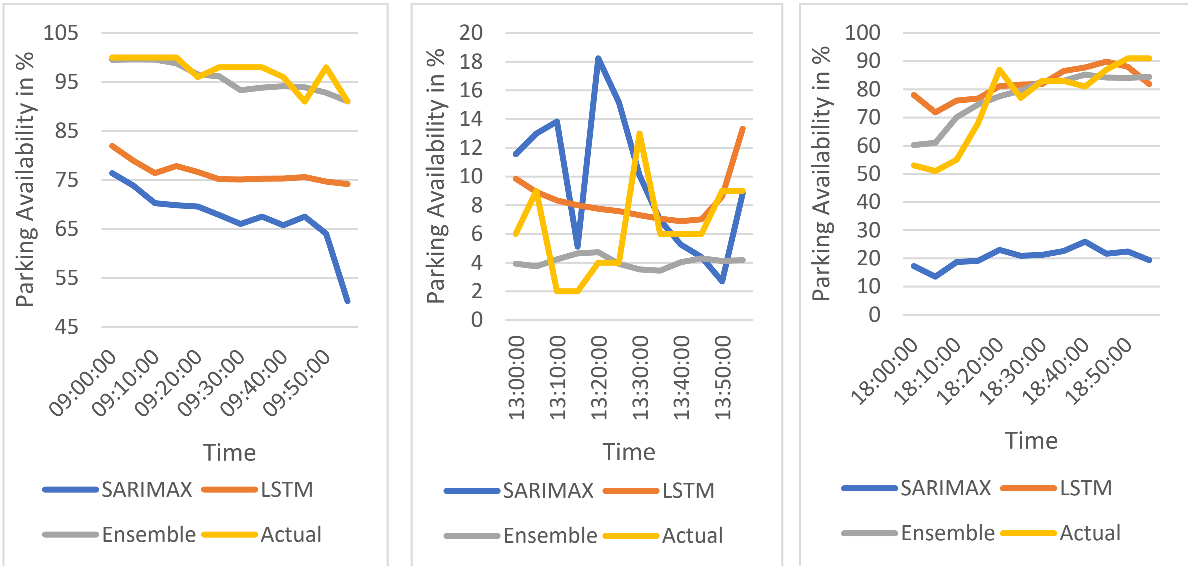
$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \tag{2}$$

4. Results and Discussion

The parking occupancy data collected from an open parking lot were evaluated using LSTM, SARIMAX, and Ensemble Methods. The sample prediction values of the three methods for weekday and weekend at multiple periods, namely 09:00, 13:00, and 18:00 are illustrated in this section. The predicted values are compared with the actual values for illustration purposes. These algorithms were trained and evaluated including and excluding visitor trend data.

4.1 Prediction with visitor trend data

Figure 4 illustrates prediction methods forecast at multiple periods, namely 09:00, 13:00, and 18:00 during a weekend. The ensemble-based method provided better short-term predictions than all other methods. As illustrated in Figure 4(b) and 4(c), the LSTM provided better predictions at 13:00 and 18:00, compared to SARIMAX. The sudden change in parking availability was learned better by Ensemble and the LSTM method. However, as shown in Figure 4(a) at 09:00, LSTM produced high errors. The SARIMAX method only produced better short-term predictions at 13:00, performing better short-term predictions when there is no sudden change of parking availability.



(a) Predictions with visitor trend data at 9 AM on weekend

(b) Predictions with visitor trend data at 1 PM on weekend

(c) Predictions with visitor trend data at 6 PM on weekend

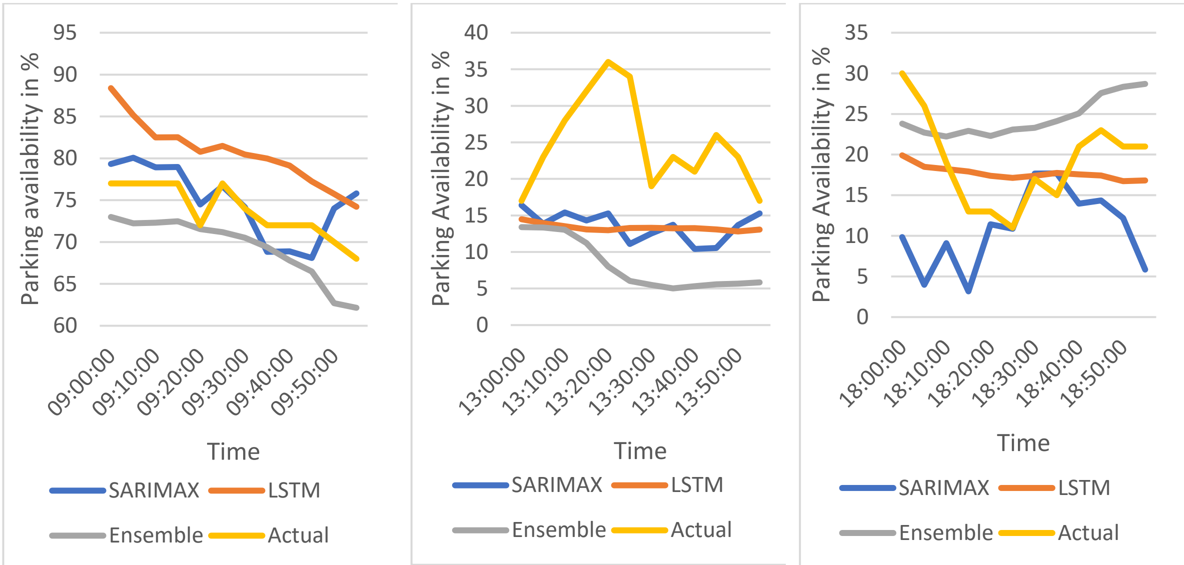
Figure 4: Prediction plot of methods with visitor trend data on the weekend

As shown in Table 4, the Ensemble-based method provided better predictions for multiple periods on a Weekend with lower MAE and RMSE values. The total dataset consists of 3 weeks of data, where only 5 days representing a weekend were utilized for training the prediction methods. Despite, the limited volume of data available for weekends, the Ensemble-based method produced better predictions. LSTM also performed better with fewer errors when compared to SARIMAX.

Table 4. Evaluation of methods with visitor trend data on weekend

	Weekend at 09:00		Weekend at 13:00		Weekend at 18:00	
Method	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**
LSTM	20.77	20.93	3.08	3.79	9.37	12.22
SARIMAX	29.79	30.13	5.22	6.93	55.1	56.44
Ensemble	1.97	2.63	3.2	4	5.9	7.2

The overall performance of the three prediction methods was better for a Weekday, as shown in Figure 5(a) and 5(c). However, when there is a sudden change in the rate of parking availability at 13:00 as shown in Figure 5(b), all three methods were unable to predict that change. Within 10 minutes, 10% of empty parking spaces were occupied. This random change in parking availability is expected for an open parking lot. However, due to limited data available for training and a lack of other potential exogenous variables, the sudden change in parking availability was not captured.



(a) Predictions with visitor trend data at 9 AM on weekday

(b) Predictions with visitor trend data at 1 PM on weekday

(c) Parking predictions with visitor trend data at 6 PM on weekday

Figure 5. Prediction of parking availability with visitor trend data on Weekday

As shown in Table 5, SARIMAX produced lower MAE and RMSE values than other methods at 09:00 on a weekday. However, all three methods produced high errors at 13:00. LSTM produced fewer errors in predicting at 18:00. No method produced lower errors for all the mentioned periods. Ensemble-based method predicts values based on aggregation and there are not enough historical data to predict the change in parking availability during peak hours. LSTM would have produced better predictions with more training data.

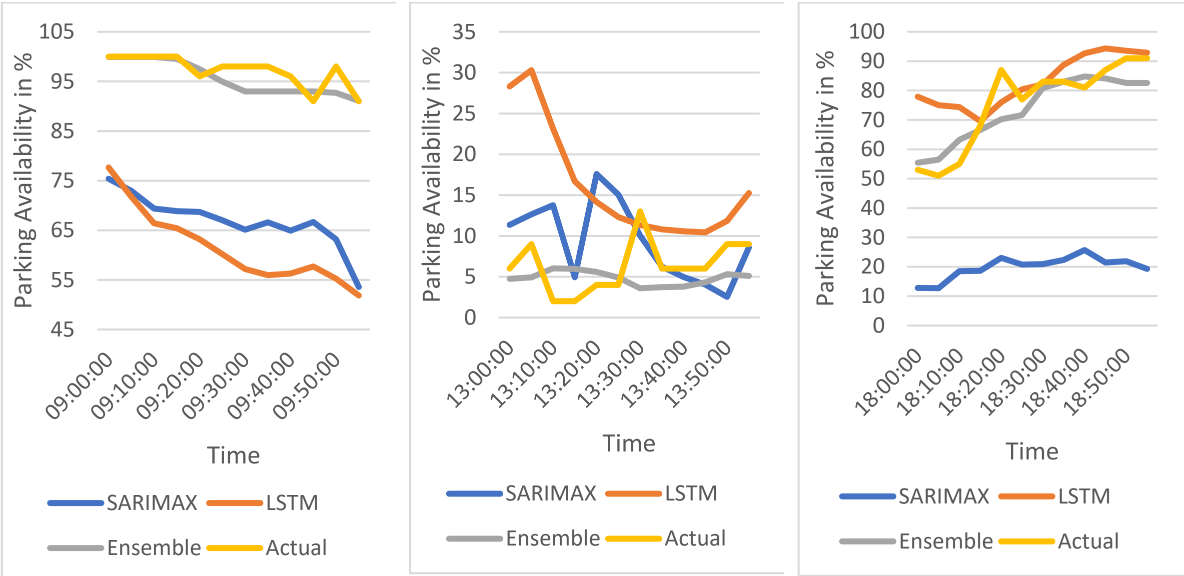
Table 5. Evaluation of methods with visitor trend data on Weekday

	Weekday at 09:00		Weekday at 13:00		Weekday at 18:00	
Method	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**
LSTM	6.88	7.12	11.58	13.19	4.53	5.21
SARIMAX	2.86	3.42	11.37	13.16	8.89	11.28
Ensemble	4.43	4.74	16.75	18	6.92	7.43

4.2 Prediction without visitor trend data

In this section, the predictions methods were evaluated without the use of visitor trend data for multiple periods mentioned in Section 4.1.

Figure 6 illustrates the prediction values without the use of visitor trend data on a Weekend. As illustrated in Figure 6(a), Ensemble-based method produced better prediction results at 09:00 and LSTM and SARIMAX produce predictions with high errors. As shown in Figure 6(b), all three methods provided predictions with lower errors. However, at 18:00, as illustrated in Figure 6(c), SARIMAX predictions produced high errors. This performance is similar to the performance with visitor trend data. Therefore, there is no change in performance of SARIMAX and Ensemble without the visitor trend data.



(a) Predictions without visitor trend data at 9 AM on weekend

(b) Predictions without visitor trend data at 1 PM on weekend

(c) Predictions without visitor trend data at 6 PM on weekend

Figure 6. Prediction plot of methods without visitor trend data for weekend

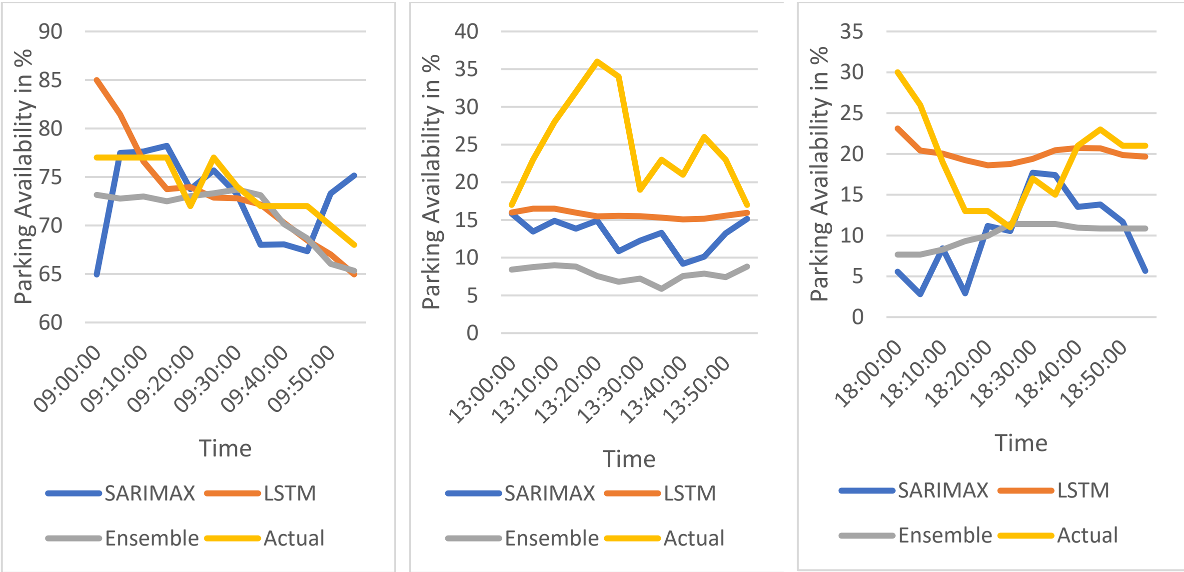
As shown in Table 6, the LSTM performed slightly better with visitor trend data when compared to without visitor trend data. The MAE and RMSE values were higher for LSTM without the use of visitor trend data. However, SARIMAX and Ensemble-based methods provided similar MAE and RMSE values with and without visitor trend data.

Table 6. Evaluation of algorithms without visitor trend data on Weekend

	Weekend at 09:00		**Weekend at 13:00**		**Weekend at 18:00**	
Method	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**
LSTM	35.6	36	10.2	12.58	9.53	12.71
SARIMAX	30.3	30.53	5.12	6.78	55.76	56.95
Ensemble	2.13	2.92	3.2	3.9	5.44	6.9

As illustrated in Figure 7, the predictions of the Ensemble-based method and SARIMAX were similar with visitor trend data. Predictions of LSTM were the only predictions affected when excluding the visitor trend data. Contrary to the performance of LSTM on a weekend, the predictions slightly improved over a weekday without the use of visitor trend data as shown in Figure 7(a) and 7(c). The variations in parking availability over a weekday were better learned by the LSTM without the use of visitor trend data. As illustrated in Figure 7(c),at 13:00, the sudden change in parking availability was not captured by any prediction methods without visitor trend data as well. As shown in Figure7(c), at

18:00, LSTM performed better, compared to other methods. SARIMAX still maintained lower parking availability at 18:00 which is the reason for larger errors.



(a) Predictions without visitor trend data at 9 AM on weekday

(b) Predictions without visitor trend data at 1 PM on weekday

(c) Predictions without visitor trend data at 6 PM on weekday

Figure 7. Prediction plot of methods without visitor trend data for Weekday

As shown in Table 7, LSTM provided better predictions with lower MAE and RMSE values compared to the Ensemble-based method and SARIMAX. The LSTM predictions were slightly improved for multiple periods on a Weekday when compared to predictions with visitor trend data. For methods such as SARIMAX and the Ensemble-based method, predictions had no impact with or without visitor trend data.

Table 7. Evaluation of algorithms without visitor trend data on Weekday

	Weekday at 09:00		**Weekday at 13:00**		**Weekday at 18:00**	
Method	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**
LSTM	2.9	3.54	9.2	11	3.84	4.6
SARIMAX	3.44	4.73	11.82	13.55	9.59	12.32
Ensemble	2.88	3.1	17	18	9.1	11

Despite the less volume of training data, the Ensemble-based method and LSTM learned the patterns more efficiently similar to results observed in (Fan et al., 2020, Tekouabou et al., 2020). Since there is a similarity in parking availability over weekends, the Ensemble-based method performed well, despite there being less volume of data. The Ensemble-based method utilizes multiple samples of the training dataset to generate aggregated values from decision trees and random forest, which is the reason for the improved prediction. But when there are quick changes in parking availability, which are not based on previous data, then the prediction methods would produce high errors in their predictions. However, this can be addressed by adding relevant exogenous or endogenous variables which affect this behavior. The LSTM and SARIMAX methods predictions are based on previous values. However, due to the usage of a memory cell in LSTM, it understood the time series data efficiently compared to SARIMAX. SARIMAX produced better prediction in (Fokker et al., 2021). However, in this study, due to less volume of data, SARIMAX produced high errors when there are sudden changes in the rate of parking availability. By increasing the volume of the dataset, the performance of the methods can further be improved. Another reason for the poor performance of SARIMAX is due to lack of parking time information.

5. Conclusion, limitations, and future work

In this paper, short-term predictions of parking availability with low volume of data were evaluated using LSTM, SARIMAX, and Ensemble-based methods. Despite trained with low-volume of data, the Ensemble-based method and LSTM produced short-term predictions with lower errors compared to SARIMAX. Abrupt changes in parking availability were not captured by any evaluated prediction methods. The use of visitor trend data only improved LSTM performance on a weekend, while SARIMAX and Ensemble-based methods were not affected by it. The results from this study are also applicable to other similar regions of open parking lots that are near to the location of interest.

The regions of the open parking lot which are near the location of interest are in high demand and usage of these prediction methods facilitated in providing a short-term prediction of parking availability. If the orientation of the camera is pointed towards the front or rear end of the vehicles, detection algorithms can be utilized to collect vehicle occupancy data. The ROI is near the shopping center which is in high demand. However, there are other regions in the parking lot which were not covered due to the limited view of the thermal camera.

Simulated data can be utilized to further improve the performance of prediction. The training dataset contains occupancy data from weekdays, weekends, holidays, and rain or snow conditions. There can be other scenarios, such as special events, severe weather conditions, which were not covered in this study. Adding relevant predictors can facilitate in further improving the accuracy of predictions. Duration of parking can also be added as a variable to improve the performance of algorithms. The parking lot can be divided into clusters to provide predictions as the rate of parking occupancy with these clusters can vary based on its proximity to the location of interest.

6. References

ASGHAR, M. N., KANWAL, N., LEE, B., FLEURY, M., HERBST, M. & QIAO, Y. 2019. Visual surveillance within the EU General Data Protection Regulation: A technology perspective. *IEEE Access,* 7**,** 111709-111726.

AWAN, F. M., SALEEM, Y., MINERVA, R. & CRESPI, N. 2020. A comparative analysis of machine/deep learning models for parking space availability prediction. *Sensors,* 20**,** 322.

BALMER, M., WEIBEL, R. & HUANG, H. 2021. Value of incorporating geospatial information into the prediction of on-street parking occupancy–A case study. *Geo-spatial Information Science,* 24**,** 438-457.

BASU, P. & LITTLE, T. D. Wireless ad hoc discovery of parking meters. MobiSys Workshop on Applications of Mobile Embedded Systems (WAMES'04), 2004.

CAICEDO, F. 2010. Real-time parking information management to reduce search time, vehicle displacement and emissions. *Transportation Research Part D: Transport and Environment,* 15**,** 228-234.

CAICEDO, F., BLAZQUEZ, C. & MIRANDA, P. 2012. Prediction of parking space availability in real time. *Expert Systems with applications,* 39**,** 7281-7290.

COOLS, M., MOONS, E. & WETS, G. 2009. Investigating the variability in daily traffic counts through use of ARIMAX and SARIMAX models: assessing the effect of holidays on two site locations. *Transportation research record,* 2136**,** 57-66.

FABUSUYI, T., HAMPSHIRE, R. C., HILL, V. A. & SASANUMA, K. 2014. Decision analytics for parking availability in downtown Pittsburgh. *Interfaces,* 44**,** 286-299.

FAN, J., HU, Q., XU, Y. & TANG, Z. 2020. Predicting the Vacant Parking Space Availability: an LSTM Approach. *IEEE Intelligent Transportation Systems Magazine*.

FıŞKıN, C. S. & CERIT, A. G. Forecasting Domestic Shipping Demand of Cement: Comparison of SARIMAX, ANN and Hybrid SARIMAX-ANN. 2019 4th International Conference on Computer Science and Engineering (UBMK), 2019. IEEE, 68-72.

FOKKER, E. S., KOCH, T., VAN LEEUWEN, M. & DUGUNDJI, E. R. 2021. Short-Term Forecasting of Off-Street Parking Occupancy. *Transportation Research Record*, 03611981211036373.

GHOSAL, S. S., BANI, A., AMROUSS, A. & EL HALLAOUI, I. A deep learning approach to predict parking occupancy using cluster augmented learning method. 2019 International Conference on Data Mining Workshops (ICDMW), 2019. IEEE, 581-586.

HOCHREITER, S. & SCHMIDHUBER, J. 1997. Long short-term memory. *Neural computation,* 9**,** 1735-1780.

IONITA, A., POMP, A., COCHEZ, M., MEISEN, T. & DECKER, S. Where to park? predicting free parking spots in unmonitored city areas. Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics, 2018. 1-12.

PAIDI, V., FLEYEH, H., HÅKANSSON, J. & NYBERG, R. G. 2018. Smart parking sensors, technologies and applications for open parking lots: a review. *IET Intelligent Transport Systems,* 12**,** 735-741.

PAIDI, V., HÅKANSSON, J., FLEYEH, H. & NYBERG, R. G. 2022. CO2 Emissions Induced by Vehicles Cruising for Empty Parking Spaces in an Open Parking Lot. *Sustainability,* 14**,** 3742.

PAIDI, V. H., JOHAN FLEYEH, HASAN AND ROGER, NYBERG 2021. CO2 emissions induced by vehicles cruising for empty parking spaces in an open parking lot.

POZO, R. F., GONZÁLEZ, A. B. R., WILBY, M. R., DÍAZ, J. J. V. & MATESANZ, M. V. 2021. Prediction of on-street parking level of service based on random undersampling decision trees. *IEEE Transactions on Intelligent Transportation Systems*.

PROVOOST, J. C., KAMILARIS, A., WISMANS, L. J., VAN DER DRIFT, S. J. & VAN KEULEN, M. 2020. Predicting parking occupancy via machine learning in the web of things. *Internet of Things,* 12**,** 100301.

RAJABIOUN, T., FOSTER, B. & IOANNOU, P. Intelligent parking assist. 21st Mediterranean Conference on Control and Automation, 2013. IEEE, 1156-1161.

RAJABIOUN, T. & IOANNOU, P. A. 2015. On-street and off-street parking availability prediction using multivariate spatiotemporal models. *IEEE Transactions on Intelligent Transportation Systems,* 16**,** 2913-2924.

RICHTER, F., DI MARTINO, S. & MATTFELD, D. C. Temporal and spatial clustering for a parking prediction service. 2014 IEEE 26th International Conference on Tools with Artificial Intelligence, 2014. IEEE, 278-282.

SHAO, W., ZHANG, Y., GUO, B., QIN, K., CHAN, J. & SALIM, F. D. Parking availability prediction with long short term memory model. International Conference on Green, Pervasive, and Cloud Computing, 2018. Springer, 124-137.

SHOUP, D. 2018. Free parking or free markets. *Parking and the City.* Routledge.

SONG, J., ZHAO, C., LIN, T., LI, X. & PRISHCHEPOV, A. V. 2019. Spatio-temporal patterns of traffic-related air pollutant emissions in different urban functional zones estimated by real-time video and deep learning technique. *Journal of Cleaner Production,* 238**,** 117881.

TEKOUABOU, S. C. K., CHERIF, W. & SILKAN, H. 2020. Improving parking availability prediction in smart cities with IoT and ensemble-based model. *Journal of King Saud University-Computer and Information Sciences*.

TIEDEMANN, T., VÖGELE, T., KRELL, M. M., METZEN, J. H. & KIRCHNER, F. Concept of a data thread based parking space occupancy prediction in a berlin pilot region. Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015.

VAN DER WAERDEN, P., TIMMERMANS, H. & BARZEELE, P. 2011. Car drivers' preferences regarding location and contents of parking guidance systems: stated choice approach. *Transportation research record,* 2245**,** 63-69.

ZHAO, Z. & ZHANG, Y. 2020. A comparative study of parking occupancy prediction methods considering parking type and parking scale. *Journal of Advanced Transportation,* 2020.