# Predicting Parking Space Availability in Paderborn Methodology Selection

Jan Lippert <ljan@mail.upb.de>

## I. INTRODUCTION

In urban areas, space and as such parking space is limited. Car owners often have to drive around to available parking spots. This leads to an unnecessary waste of time and additional air pollution.

According to multiple studies, the availability of parking data reduces the search time for a parking spot [1][2]. However, most systems only provide the current number free parking spots. In rush hours, this information can quickly get outdated and lead to driving around to find available parking spaces.

The goal of this project is to predict the available parking space in advance. Because of the limitation outlined in section **??**, we will focus on the covered car park "Libori-Galerie". This car park is also the most interesting as it is directly connected to a local shopping center. It is opened from 06:00 am to 02:00 am.

## II. DATA SOURCES

The main data source is the actual parking data which is available online. In addition there are other possible features that may influence parking space availability. Since taking all features into account is not possible, this project will solely focus on the parking data in relation to time.

*a) Parking Usage Live Data:* The homepage of the ASP Paderborn[1] lists the city managed parking areas and the number of currently free parking spots. The data is available on https://www.paderborn.de/microsite/asp/parken_in_der_city/freie_Parkplaetze_neu.php. A more minimal website is available at https://www4.paderborn.de/ParkInfoASP/default.aspx.

The website displays 4 attributes: name, type, capacity, and available spots[2]. After crawling, the parking information will be annotated with the crawling time. The latter will be split into multiple attributes which will be used in the prediction: hour of day $h$, minute of hour $m$, day of week $d$, day of month $d_m$, week of month $w_m$, week of year $w_y$, and $a$[3].

The number of available parking spaces is be called $y$. With the features defined before, the training data will be a set of tuples of the form
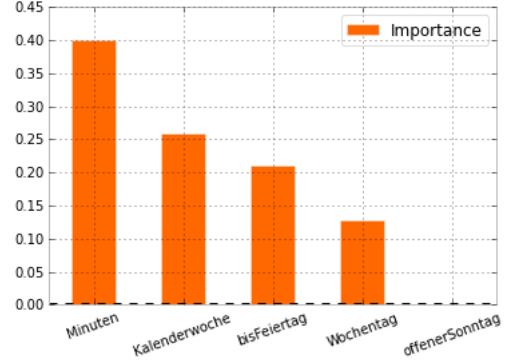
$$(h, m, d, d_m, w_m, w_y, a, y).$$



Fig. 1. Importance of Features (ParkenDD)

*b) Event Data (Optional):* As mentioned in the first report, public holidays and local events could influence the parking situation near the city center. In some cases, these events take place *on* one of the parking areas. However, as most of the big events do not happen while the course "Practical Project in Machine Learning" takes place, we will not consider them further.

Public holidays like Christmas and Easter may also influence the availability of parking space. [**?**, parkendd]odelled this in the feature "bisFeiertag", i.e. working days until public holiday. In figure 1, the relevance of the different features used by [3] is displayed[4].

The feature was relatively important in the evaluation done by [3]. Nonetheless, it will not be included as the next public holiday is Christi Himmelfahrt on 25th May 2017. Therefore there would be no possibility of verifying the importance of this feature in time.

*c) Weather Data (Optional):* Weather data may also prove to be useful for prediction as, e.g., more people may decide to take the car when it's raining. Weather data is available online, e.g. at https://openweathermap.org/api. Weather data will not be considered in the prediction to keep the model simpler.

## III. HYPOTHESIS

It is assumed, that the available parking space is dependent on time, i.e. there exists a function $f$ such that

$$f(h, m, d, d_m, w_m, w_y, a) = y.$$

---

[1]City-Managed service for waste management, city cleaning, and parking.

[2]respectively: Parksttte, –, Anzahl, Frei

[3]the year is included as a way to keep samples unique in case the project runs more than 1 year.

[4]respectively: minutes, week of year, working days until public holiday, day of week, Sunday openings

## IV. Feasibility

To keep the scope of this project small, multiple data sources will be ignored. For this project, it is assumed that the number of available parking spots solely depends on time. As described in section II, this obviously is a very simplified model. However, as shown by [3], such a simple model may be enough.

The greatest challenge in this project is to collect enough training data to make meaningful predictions. All data needs to be extracted from APS's website. Changes to the website and non-availability of the service may prove problematic. In the beginning of February, the ASP parking guiding system was dysfunctional and did not show useful data online.

The next limit is the usage of heroku's free plan. This plan limits the used database to 10000 rows. When crawling every 15 minutes, this limits the amount of available data to roughly one week before heroku disables inserts into the database. The 10000 row limit can be exceeded for 7 days before inserts are disabled.

## V. Model

The main focus will be on using a regression tree to model the availability rate as this seems to be the most promising model. It was used by both [3] and [4] with great success. One appealing reason to use this model is the possibility of visualizing the prediction algorithm.

Another reason is that some of the prediction features are categoric in nature, e.g. day of week. This is more of a side-note as these features can be converted to numerical data, i.e. $0$ to $6$ instead of $monday$ to $sunday$.

As regression trees may be very inaccurate, other models like Support Vector Machines or Multivariate Linear Regression may be investigated. The accuracy of the Regression Tree used by [3] was most likely achieved by limiting the predicted labels to buckets of 10%s, ie. 0%-10% free, 10%-20% free, etc.

## VI. Verification

To verify the accuracy of the predictions, the prediction system will be queried in regular intervals. The predictions $y_i^*$ and the actual values $y_i$ are then compared via mean-absolute error:

$$MAE := mean(|y_i^* - y_i|).$$

### References

[1] Y. Asakura and M. Kashiwadani, "Effects of parking availability information on system performance: a simulation model approach," in *Vehicle Navigation and Information Systems Conference, 1994. Proceedings., 1994*, pp. 251–254, Aug 1994.

[2] F. Caicedo, "Real-time parking information management to reduce search time, vehicle displacement and emissions," *Transportation Research Part D: Transport and Environment*, vol. 15, no. 4, pp. 228 – 234, 2010.

[3] P. Balzer, "Vorhersage der parkhausbelegung mit offenen daten," Mar. 2015. Available: http://mechlab-engineering.de/2015/03/vorhersage-der-parkhausbelegung-mit-offenen-daten/ [Accessed: 24.01.2017].

[4] Y. Zheng, S. Rajasegarar, and 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)*, pp. 1–6, April 2015.