Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 201 (2022) 40–47
Procedia Computer Science
www.elsevier.com/locate/procedia

The 13th International Conference on Ambient Systems, Networks and Technologies (ANT)
March 22-25, 2022, Porto, Portugal

Predicting truck parking occupancy using machine learning

Stefani Slavova, Jean Paul Sebastian Piest*, Wouter van Heeswijk

University of Twente, Drienerlolaan 5, 7522 NB Enschede, the Netherlands

Abstract

The logistics industry faces an increasing shortage of truck parking spots. This results in illegal parking or fatigued driving with hazardous consequences for traffic safety, as truck drivers have no insight into future availability of parking spots. Accurate short-term predictions of parking lot occupation are required to aid drivers in planning their routes and rest stops. To obtain such predictions, this research compares a variety of machine learning algorithms, concluding that decision trees are most suitable for real-time application. The model is trained on real-world data containing 1.5 years of truck parking measurements, obtained from a truck parking in Deventer, the Netherlands. We find that – contrasting to car parking, which is influenced by factors such as the weather – a model using only temporal features and historical occupancy yields the best results. For one-hour ahead predictions, we obtain an RMSE of 0.0029, with a training time of 4 seconds and predictions being sufficiently fast for real-time deployment. The main contributions of this research are (i) a machine learning approach for predicting truck parking occupation, (ii) insights into relevant predictive features, and (iii) a case study. From a practical perspective, we propose an architecture for a dynamic prediction tool, which can be used by truck drivers, parking managers and road authorities to improve truck parking utilization. Future research can build upon the machine learning approach and use the prediction model for other truck parking areas.

Keywords: truck parking; machine learning; predicting parking occupancy; case study.

1. Introduction

Road freight comprises 53.4% of the intra-European trade and logistics industry, followed by maritime and rail transport, accounting for 29.6% and 12.3% respectively [1]. Every day, goods worth billions of euros are transported

* Corresponding author. Tel.: +31 5 34895981.
E-mail address: j.p.s.piest@utwente.nl

10.1016/j.procs.2022.03.008

via the Trans European Road Network – which serves as the backbone of trade and commerce on the European continent – underlining the economic dependence on trucks. Currently, about 6.5 million trucks circulate throughout the European Union (EU) [2], necessitating sufficient supporting infrastructure and -legislation.

The number of operating trucks keeps growing steadily [3], leading to parking capacity problems, nuisance, and unsafe traffic. According to a European Commission (EC) study [4], 83% of truck drivers believe that Europe lacks sufficient safe and secure truck parking areas. The study revealed a remarkable shortage of 400,000 secure parking spots. Drivers unable to find a suitable parking space might choose to either park somewhere illegally or continue driving in a state of fatigue. Highway access ramps and emergency lanes are common unauthorized parking locations, creating dangerous traffic conditions. Continuing driving without rest imposes safety hazards. Several studies have linked fatigued drivers to an increased accident risk [5]; a 2014 study revealed that 21% of all fatal accidents involve fatigued drivers [6]. Therefore, EC regulations [7] impose driving time limits and set rest periods.

The parking demands during the night and on the weekends frequently exceed the parking capacity. However, while some parking areas are completely full, others may have free spots. A key reason for this imbalance is the lack of information about parking areas, their facilities, and the availability of free parking spots [8]. As road freight transport is expected to grow, optimizing the existing capacity of truck parking areas by providing travel information services is listed as a top priority by the EC [9].

In this paper, which is based on a graduation thesis [10], we propose a machine learning approach for predicting the occupancy of truck parking spots, using historical data to dynamically predict the occupancy rates of truck parking locations. Such predictions could aid drivers in better planning their next stop. To identify a suitable prediction model, we test and compare several state-of-the-art machine learning algorithms in a real-life setting at A1 Truck Parking, following the CRoss Industry Standard Process for Data Mining (CRISP-DM) [11].

The contributions of the study are the following:

- It fills an important literature gap, as extant literature focuses almost exclusively on car parking occupancy prediction and consistent research about truck parking is missing;
- It proves the applicability of machine learning algorithms to forecast the occupancy of truck parking areas using historical data. Various sets of predictive features are designed and evaluated to improve model accuracy;
- It addresses the problem of suboptimal parking utilization by proposing a dynamic information service about the current availability of parking places.

The rest of the paper is organized as follows. Section 2 analyses related literature. Section 3 presents the case study results, following the CRISP-DM cycle. Section 4 discusses the key findings. Section 5 concludes the paper. Section 6 discusses future work.

2. Related work

Our literature review on truck parking occupancy prediction reveals clear knowledge gaps; few studies have explored the topic. Tavafoghi et al. [12] present a queuing model approach for probabilistic prediction of future parking occupation, assuming that a parking lot has infinitely many parking spots. Considering the truck parking shortage observed in real life, this inordinate assumption limits the practical applicability of the model. Sadek et al. [13] propose a Fourier transformation approach, which successfully predicts the occupancy for any time within the present day. However, their data is highly accurate due to the truck stop's inventory system, where drivers check in with staff. As this is not the case at most conventional truck stops – where data often suffers from errors and noise – the authors acknowledge limited applicability towards other parking areas, until sensory technology improves.

While truck parking occupancy prediction is rarely addressed in the literature, many studies investigate urban parking occupancy prediction, such as car parking. Specifically, machine learning approaches have demonstrated effectiveness in forecasting occupation. We examined nine papers that deploy machine learning prediction, synthesizing both input features and utilized techniques. The utilized sets of features are summarized in Table 1. Additional features include location and number of parking lots. When it comes to the choice of learning algorithms, studies explored a variety of state-of-the-art machine learning techniques. Decision trees receive much consideration due to their transparency, elucidating the correlation between features, and their low computational demands. Zheng et al. [16] compare decision tree performance to support vector machines and neural networks, demonstrating that decision trees outperform the alternatives. The suitability of decision trees is corroborated by Reinstadler et al. [17] and Fabusuyi et al. [18]. Next, studies positively regard the random forest algorithm. Chawathe [19] and Kim and Koshizuka [20] test multiple regression techniques, demonstrating that random forests achieve the highest accuracy. Finally, neural networks appear to generate promising results. Provoost et al. [14] show that a feed-forward neural network outperforms random forest. Vlahogianni et al. [21] build a parking availability prediction system based on neural networks and emphasize their ability to capture the temporal evolution of parking occupancy. Pflügler et al. [22] take a similar stand, underlining the suitability of neural networks to predict occurrences when the underlying relationships are not fully understood. Nevertheless, the 'black box' nature of neural networks restrains stakeholders from understanding the precise contribution of features, limiting applicability [14].

Table 1. Matrix of independent features used in parking occupancy prediction models; adapted from [10].

Article	Time of the day	Weekday	Historical occupancy	Traffic flow	Rainfall	Temperature	Holiday	Event	Other
Provoost et al. [14]	X	X	X	X	X	X			
Chen [15]	X	X						X	X
Zheng et al. [16]	X	X	X						
Reinstadler et al. [17]	X				X	X	X	X	
Fabusuyi et al. [18]	X	X	X		X			X	
Chawathe [19]	X	X							
Kim & Koshizuka [20]	X	X	X						X
Vlahogianni et al. [21]	X	X	X						
Pflügler et al. [22]	X	X		X	X	X	X	X	X

Despite the variety of approaches, deploying machine learning techniques to predict car parking occupancy produces promising results. Considering the limited research into truck parking occupancy prediction, especially by means of machine learning, we aim to reduce the identified literature gap. As not much is known about the features influencing truck parking behaviour, we base our design on urban parking occupancy prediction. Finally, we offer practical insights based on case study research at a representative truck parking.

3. Case study: A1 Truck Parking

The case study is conducted at A1 Truck Parking [23] in Deventer, the Netherlands, on behalf of the Province of Overijssel. The study is part of a long-term program focused on improving the truck parking situation in the region. To support the efficient use of existing truck parking infrastructure, we investigate the possibility of developing an information system that is capable of embedding machine learning algorithms to predict the occupancy of truck parking areas. The main aim is to identify a suitable machine learning algorithm and a set of input features to build a robust prediction model. The case study is structured according to the steps of the CRISP-DM methodology.

3.1. Business understanding

We first identify involved stakeholders and expected benefits from implementing a predictive information system. Based on interviews with the Province of Overijssel, the main stakeholders are truck drivers, goods owners, road haulage companies, road users, parking infrastructure owners, governmental bodies, and IT system developers.

Anticipated benefits of the dynamic information service include:

- More effective route planning and thus, fewer parking-related issues and compliance with statutory breaks;
- More trucks parked at safe and secure locations, increasing safety and improved wellbeing for truck drivers;
- Fewer cargo crimes and fewer related costs or lost revenues for goods owners and road haulage companies;

- Fewer accidents as a result of roadside parking or fatigued truck drivers, leading to safer traffic;
- An improved traffic flow due to fewer wandering trucks;
- Less nuisance caused by illegal truck parking for business park managers and nearby communities;
- Increased revenues for parking infrastructure owners due to higher utilization of services and products offered;
- Lower costs for the construction and maintenance of new parking infrastructure due to better utilization.

3.2. Data understanding

Following context analysis and observations from related works, available data were gathered to perform the case study. Truck parking data spanning over 1.5 years (January 2020 - June 2021) was provided by A1 Truck Parking Deventer, which has a capacity of 100 parking places. Data consists of electronic toll transactions (timestamps of each in- and outgoing vehicle), from which occupancy rates are derived. Furthermore, we obtained hourly weather data (temperature, rainfall) and Dutch and German holidays from public online sources [24,25,26].

As shown in Table 1, time features are among the most prominent predictors of parking occupancy. To confirm this, we investigated the changes in the occupancy rate according to the hour of the day and day of the week. The results, presented in Figure 1, indicate both hourly and daily fluctuations. The parking demand shows spikes during the night and at the weekend, which is in line with the presented problem description.

Fig. 1. (a) Hourly variations; (b) Daily variations.; adapted from [10]. Clear temporal patterns may be discerned both intra-day and inter-day.

3.3. Data preparation

The truck parking dataset contained three weeks of missing data (3.7%), due to a temporary closure of the parking entrance for construction work. The parking area was operating as usual during this period. Therefore, we replaced the missing values by the means of an imputation method. To reflect the hourly and daily seasonality of the data, we used a custom averaging technique – calculating the mean occupancy rates per hour of the day for each weekday and mapping them onto the missing values.

We established the dataset for modelling by combining the truck parking, weather, and holiday datasets. From the timestamps, we engineered temporal features (i.e., time of the day, day of the week). All data were resampled to fit one-hour intervals. To provide the model with recent occupancy rates, we created a lookback window of eight hours. As shown in Figure 1(a), occupancy rates do not change significantly within a few hours, indicating the need for a longer lookback window. Based on a pairwise correlation feature selection process, we preserved the historical occupancy from the previous hour, four hours ago, and seven hours ago, due to their high correlation with the dependent variable and only moderate correlation with each other.

3.4. Modelling

To select an appropriate machine learning technique for modelling, we used RapidMiner [27]; a data science software platform that enables simultaneous comparison of multiple algorithms based on performance and runtime.

For our problem – using the Root Mean Squared Error (RMSE) and runtime as performance metrics – RapidMiner recommended the following algorithms: generalized linear model, artificial neural network, decision tree, random forest, gradient boosted trees, and support vector machine. As shown in Table 2, no technique scores best on both metrics (i.e., error and runtime). Due to its lowest runtime and third-best RMSE, we selected the decision tree algorithm for the experiment. The decision tree algorithm was also positively regarded in related works.

Table 2. Results of comparing multiple machine learning algorithms; adapted from [10]. Decision trees display the lowest runtime and high accuracy.

Machine learning algorithm	RMSE	Standard deviation	Runtime
Generalized linear model	0.02	0.002	7 sec
Artificial neural network	0.018	0.001	30 sec
Decision tree	**0.02**	**0.001**	**4 sec**
Random forest	0.022	0.001	52 sec
Gradient boosted trees	0.018	0.002	1 min 3 sec
Support vector machine	0.02	0.002	29 min 18 sec

For the model development, the Python library Scikit-learn was utilized. The complete dataset was divided into training (80%) and testing (20%) sets. Due to the sequential nature of time series, the temporal order of the observed values must be preserved [28]. This ensures that future observations are not used in constructing the predictions, allowing proper model validation. In addition, we used a variant of 5-fold cross-validation for hyperparameter tuning, called time series split and described as “evaluation on a rolling forecasting origin” [29]. The maximum depth and minimum sample split of the decision tree were tuned simultaneously by applying a grid search.

For the feature selection we build two candidate models, both visualized in Figure 2. The first candidate model includes the preselected input features, as observed in related work about urban parking occupancy prediction (see Table 1). For the second candidate model, we applied a feature elimination strategy. In our experiments, we sequentially excluded a category (or multiple categories) of features to determine their importance for the prediction. The features were clustered based on their data source and nature (e.g., temperature and rainfall; holidays in the Netherlands and in Germany). The experiment yielded a second candidate model, with a simpler configuration, which slightly outperforms the first candidate model when tested on the training set. The experiments suggest that truck parking behaviour is not heavily influenced by meteorological conditions and holidays. For accurate predictions, utilizing time features and historical occupancy suffices and actually slightly improves performance.

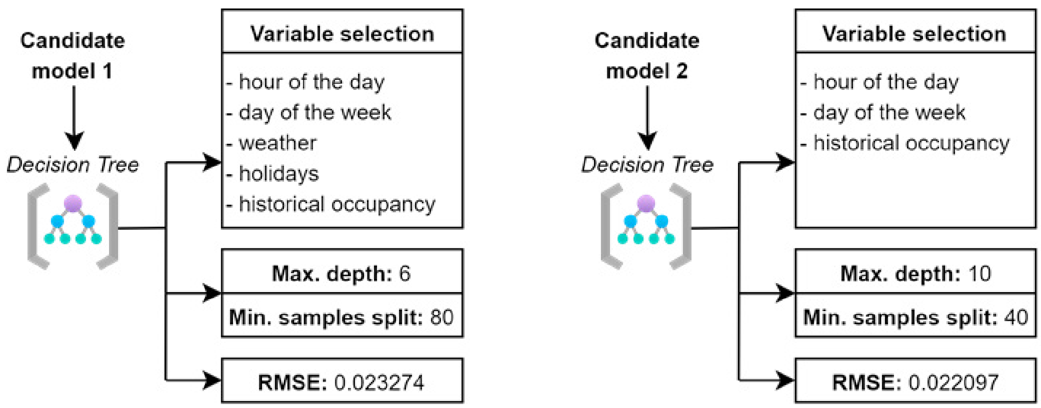

Fig. 2. (a) Candidate model 1 (preselected features); (b) Candidate model 2 (feature elimination strategy); adapted from [10].

3.5. Evaluation

To further compare performance of both candidate models using the test set, we selected the RMSE, Mean Absolute Error (MAE), R^2 (coefficient of determination) and adjusted R^2. The results (see Table 3) show that the second candidate model outperforms the first on every metric, with lower prediction errors and higher correlation coefficients. Therefore, we selected the second decision tree as our prediction model. This model configuration has an additional advantage regarding its ease of implementation, as all input features can be derived from a singular data source; the model requires fewer resources and development efforts. Additionally, it is less computationally demanding; for a real-time predictive system, this translates into an enhanced user experience.

Table 3. Evaluation of both candidate models on the test set; adapted from [10].

Metric	Candidate model 1	Candidate model 2
RMSE	0.024068	0.022948
MAE	0.014649	0.013330
R^2	0.923590	0.930538
Adjusted R^2	0.923289	0.930375

3.6. Deployment

Next, we discuss the information system requirements to address deployment issues. To deploy our prediction model in practice, we propose an integrated dynamic information system geared to the needs of truck drivers. The system enables better route planning based on occupation of nearby parking lots. The system architecture (Figure 3) integrates data collection, pre-processing, storage, predictions, evaluation, and communication.

A key challenge for implementing intelligent transportation services is gathering the data from all actors involved. Our solution incorporates a shared cloud, providing hourly data from each truck parking's electronic toll system. Adequate security policies are ensured by an identity and access management system. Parking lot occupation is updated on an hourly basis by measuring ingoing- and outgoing trucks. The time-dependent features and historical occupancy are simultaneously generated and stored in the database. Using the most recent input data, the decision tree predicts parking occupancy for the next hour. Based on the differences between observations and predictions, the decision tree is periodically updated in an online fashion. An Application Programming Interface (API) links third parties to the system. The system informs truck drivers via a mobile application, whereas road authorities can incorporate truck parking occupancy in Dynamic Route Information Panels (DRIPs).

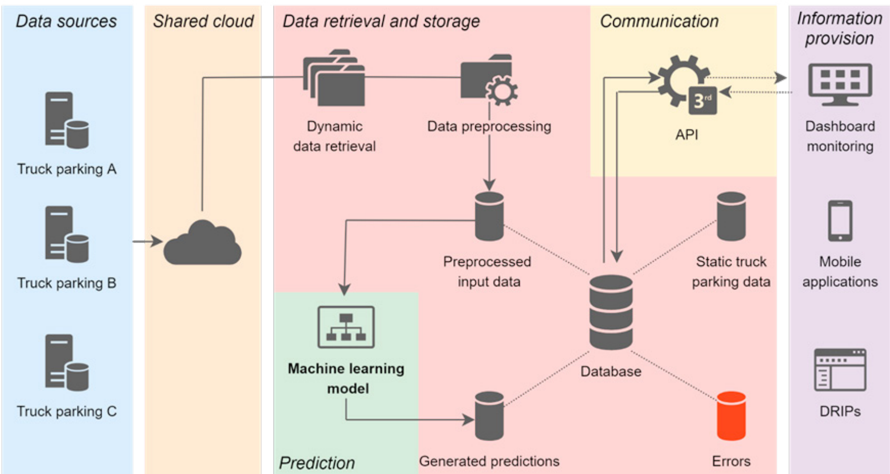

Fig. 3. Conceptual architectural design of an integrated predictive system; adapted from [10].

4. Discussion

Prior research indicated the potential of machine learning models to accurately predict urban parking occupancies. Despite the urgency of truck parking issues, machine learning has received limited research attention in that domain. Practitioners mainly apply traditional statistical approaches, which show limited suitability to support real-time occupancy prediction. Furthermore, existing models do not integrate and examine external factors affecting truck parking behaviour, such as time of the day, day of the week, weather, or holidays.

Our work contributes to existing literature by applying machine learning to predict truck parking occupancy, exploring the integration of multi-source feature representation. Our findings indicate that, in contrast to studied urban parking occupancy prediction models – where accuracy increases by adding exogenous features (e.g., holiday, weather) – additional features do not contribute to more reliable forecasting of truck parking occupation. The case study demonstrates that temporal features and occupation rates in the recent past suffice for accurate predictions. The likely explanation is that freight transport tends to be stable regardless of, e.g., weather circumstances. However, we note that the climate in the Netherlands is relatively moderate. For countries with more extreme weather conditions, adding features representing the weather might enhance the performance of the model.

5. Conclusion

This research provides new insights on solving the increasingly urgent truck parking problem, suggesting a dynamic information service supported by a machine learning prediction model. Our results demonstrate that a decision tree fed with data from a single data source (historical occupancy rates) can predict the next-hour occupancy rate with a RMSE of 0.0029. The decision tree was benchmarked against a number of other machine learning algorithms, outperforming them in terms of runtime while offering competitive accuracy.

The proposed approach brings novelty to the truck parking occupancy prediction field, deploying machine learning rather than traditional forecasting approaches. The latter generally require more domain knowledge, restricting their applicability to specific conditions and limiting their scalability over multiple parking lots. In contrast, machine learning models need more data to discern meaningful patterns but require less domain knowledge, making them easier to implement on a larger scale.

Despite the promising results, due to data limitations, we were unable to validate whether the proposed approach applies to other truck parking locations. However, the analysis suggests that there is unrealized potential in data-driven parking decisions. From our point of view, more research and innovation are needed to mitigate the increasing commercial truck parking capacity problems.

6. Future work

For follow-up research, our goal is to collect additional data from geographically dispersed parking lots to validate the approach and test the model generalization for other cities and countries. By studying various settings, we aim to further improve the prediction model by understanding which other predictive variables affect the truck parking occupancy. Furthermore, we plan to assess the model's performance in an integrated cloud environment under operational conditions. Sensor-, camera- and GPS data might be leveraged to further improve the accuracy and robustness of predictions. Finally, the proposed predictions system must be tested in real life to assess its suitability to address the truck parking problem. The aim is to utilize predictions to support parking decisions and better spread parking lot occupancy; this decision-support capability requires further exploration. The proposed integrated predictive system can be positioned as part of the logistics data sharing infrastructure that is being developed in the Netherlands, seeking alignment with the international data spaces movement in the EU.

Acknowledgements

The authors would like to thank the Province of Overijssel and A1 Truck Parking Deventer for their support.

References

[1] Eurostat. (2021). *Freight transport statistics - modal split - Statistics Explained*. From: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Freight_transport_statistics_-_modal_split#Inland_freight_transport_performance_-the_need_to_adjust_road_transport

[2] ACEA - European Automobile Manufacturers' Association. (2017, August 3). *Fact sheet: Trucks*. From ACEA: https://www.acea.be/publications/article/factsheet-trucks

[3] Eurostat. (2019). *Road freight transport statistics - Statistics Explained. April, 1–10. From:* https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Road_freight_transport_statistics

[4] European Commission. (2019). *Study on Safe and Secure Parking Places for Trucks* (Issue February).

[5] European Commission. (2015). Fatigue 2015. *European Commission, Directorate General for Transport*.

[6] Tefft, B. C. (2014). *Prevalence of motor vehicle crashes involving drowsy drivers, United States*, 2009-2013. Washington, DC: AAA Foundation for Traffic Safety.

[7] Regulation EC No 561/2006. (2006). *The harmonisation of certain social legislation relating to road transport.* European Parliament, Council of the European Union. http://data.europa.eu/eli/reg/2006/561/oj

[8] Nagy, E., & Sandor, Z. (2012). Intelligent truck parking on the Hungarian motorway network. *Pollack Periodica*, 7(2), 91–101. https://doi.org/10.1556/Pollack.7.2012.2.8

[9] Directive 2010/40/EU. (2010). *The framework for the deployment of Intelligent Transport Systems in the field of road*. European Parliament, Council of the European Union. http://data.europe.eu/eli/dir/2010/40/oj

[10] Slavova, S.I. (2021) Predicting the occupancy rates of truck parking locations: a machine learning approach. From: https://essay.utwente.nl/88346/

[11] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Rudiger, W. (2000). Crisp-Dm 1.0. *CRISP-DM Consortium, 76*

[12] Tavafoghi, H., Poolla, K., & Varaiya, P. (2019). A Queuing Approach to Parking: Modeling, Verification, and Prediction. *arXiv preprint arXiv:1908.11479*.

[13] Sadek, B. A., Martin, E. W., & Shaheen, S. A. (2020). Forecasting Truck Parking Using Fourier Transformations. *Journal of Transportation Engineering, Part A: Systems, 146(8), 05020006*.

[14] Provoost, J., Wismans, L., Van der Drift, S., Kamilaris, A., & Van Keulen, M. (2019). Short Term Prediction of Parking Area states Using Real Time Data and Machine Learning Techniques. *arXiv preprint arXiv:1911.13178*.

[15] Chen, X. (2014). Parking occupancy prediction and pattern analysis. *Dept. Comput. Sci., Stanford Univ., Stanford, CA, USA, Tech. Rep. CS229-2014*.

[16] Zheng, Y., Rajasegarar, S., & Leckie, C. (2015, April). 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). IEEE.

[17] Reinstadler, M., Braunhofer, M., Elahi, M., & Ricci, F. (2013). Predicting parking lots occupancy in Bolzano. *Academic Project, Computer Science, Free University of Bolzano Italy, Bolzano.*

[18] Fabusuyi, T., Hampshire, R. C., Hill, V. A., & Sasanuma, K. (2014). Decision analytics for parking availability in downtown Pittsburgh. *Interfaces*, 44(3), 286-299.

[19] Chawathe, S. S. (2019, October). Using Historical Data to Predict Parking Occupancy. In *2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)* (pp. 0534-0540). IEEE.

[20] Kim, K., & Koshizuka, N. (2019, October). Data-driven parking decisions: Proposal of parking availability prediction model. In *2019 IEEE 16th International Conference on Smart Cities: Improving Quality of Life Using ICT & IoT and AI (HONET-ICT)* (pp. 161-165). IEEE.

[21] Vlahogianni et al. (2016): Vlahogianni, E. I., Kepaptsoglou, K., Tsetsos, V., & Karlaftis, M. G. (2016). A real-time parking prediction system for smart cities. *Journal of Intelligent Transportation Systems*, 20(2), 192-204.

[22] Pflügler, C., Köhn, T., Schreieck, M., Wiesche, M., & Krcmar, H. (2016). Predicting the availability of parking spaces with publicly available data. *Informatik 2016*.

[23] A1 Truck Parking: https://a1truckparkingdeventer.com/

[24] Koninklijk Nederlands Meteorologisch Instituut (KNMI). (n.d.). Uurwaarden van weerstations., from https://www.daggegevens.knmi.nl/klimatologie/uurgegevens

[25] Ministerie van Algemene Zaken. (n.d.). Which days are official public holidays in the Netherlands? Government.Nl., from https://www.government.nl/topics/working-hours/question-andanswer/public-holidays-in-the-netherlands

[26] Bundesamt für Güterverkehr (BAG). (n.d.). Feiertags-Fahrverbote für Lkw in Deutschland im Jahr 2021. Bundesamt für Güterverkehr. https://www.bag.bund.de/DE/Themen/RechtsentwicklungRechtsvorschriften/Rechtsvorschriften/Strassenverkehrsrecht/LKWFahrverbote/Feiertagsfahrverbote2021/Feiertagsfahrverbote2021neu.html

[27] RapidMiner: https://rapidminer.com/

[28] Brownlee, J. (2019, August 28). *How To Backtest Machine Learning Models for Time Series Forecasting*. Machine Learning Mastery. https://machinelearningmastery.com/backtest-machine-learningmodels-time-series-forecasting/

[29] Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: an analysis and review. *International journal of forecasting*, 16(4), 437-450.