

Article

Not peer-reviewed version

Highway Rest Area Truck Parking Occupancy Prediction Using Machine Learning: A Case Study from Poland

[Artur Budzyński](#) and [Maria Cieśla](#) *

Posted Date: 15 May 2025

doi: 10.20944/preprints202505.1162.v1

Keywords: highway rest area; truck parking occupancy; machine learning; transport infrastructure; predictive modeling; highway safety



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Highway Rest Area Truck Parking Occupancy Prediction Using Machine Learning: A Case Study from Poland

Artur Budzyński ^{1,*} and Maria Cieřła ^{2,*}

¹ Department of Product Packaging Science, Institute of Quality Sciences and Product Management, Cracov University of Economics, 27 Rakowicka St., 31-510 Cracov, Poland

² Department of Transport Systems, Traffic Engineering and Logistics, Faculty of Transport and Aviation Engineering, Silesian University of Technology, 8 Krasińskiego St., 40-019 Katowice, Poland

* Correspondence: abudzyns@uek.krakow.pl (A.B); maria.ciesla@polsl.pl (M.C.)

Abstract: Highway rest areas are relevant components of road infrastructure, providing drivers with essential opportunities to rest and mitigate fatigue-related crash risks. Despite their acknowledged importance, little is known about the factors that influence their actual utilization. This study addresses this gap by applying supervised machine learning algorithms to predict hourly occupancy levels of truck parking lots at highway rest areas using a dataset collected from digital monitoring systems in Poland. The dataset includes 10,740 observations and 33 features describing infrastructural, administrative, and locational characteristics of selected rest areas in Poland. Five classification models—Decision Tree, Random Forest, Gradient Boosting, Logistic Regression, and XGBoost—were implemented and compared using standard performance metrics. Gradient Boosting emerged as the best-performing model, achieving the highest prediction accuracy and identifying key features such as the presence of fuel stations, rest area category, and facility amenities as significant predictors of occupancy. The findings highlight the potential of interpretable machine learning methods for supporting infrastructure planning, particularly in identifying underutilized or overburdened facilities. This research demonstrates a data-driven approach for analyzing rest area usage and provides practical insights for optimizing facility distribution, enhancing road safety, and informing future investments in transport infrastructure.

Keywords: highway rest area; truck parking occupancy; machine learning; transport infrastructure; predictive modeling; highway safety

1. Introduction

1.1. Background and Motivation

Highway rest areas (HRAs) are essential components of road infrastructure serving multiple functional, safety, environmental, and economic roles. Rest areas (RAs) support overall traffic efficiency by distributing stopping and resting behavior along highways in a controlled and predictable manner. They reduce the incidence of unsafe roadside parking, particularly by freight vehicles, which can otherwise obstruct emergency lanes and exit ramps [1]. Moreover, they serve as logistical hubs, sometimes integrated with weigh stations, fueling facilities, and freight management systems. The European Commission promotes “Safe and Secure Truck Parking Areas” (SSTPAs) across Trans-European Transport Networks (TEN-T), recognizing rest areas as strategic infrastructure assets [2]. These are designated areas outside the road lane that have been specially modified to accommodate the demands of drivers and travelers. They are equipped with parking spaces, sanitary, catering and accommodation infrastructure, as well as facilities for refueling and servicing the vehicle.

Motorway rest areas are providing drivers with opportunities to rest and mitigate fatigue, a significant factor contributing to road accidents. Fatigue, recognized as a critical factor in road crashes [3], impairs cognitive functions such as reaction time, attention, and decision-making, increasing the likelihood of collisions. Furthermore, research indicates that driver fatigue is insufficiently recognized and reported as a cause of road accidents [4]. The presence of well-maintained rest areas has been associated with a reduction in fatigue-related crashes, highlighting their role in enhancing road safety [5]. Therefore, understanding and improving the utilization of HRAs is crucial for developing strategies to combat driver fatigue and its associated risks.

Recent studies have further emphasized the importance of rest area availability. For instance, research indicates that fatigue-related commercial vehicle driver at-fault crashes occurred more frequently when RAs were more than 20 miles apart [6]. Segments nearer rest locations had less fatigue-related occurrences, according to [7] studies on the relationship between points proximity and fatigue-related incidents, indicating a beneficial safety impact. Additionally, a study examining the impact of limited RAs on truck driver crashes in Saskatchewan found a significant association between fatigue-related crashes and a lack of truck stops and RAs [8]. These findings underscore the need for strategically located and adequately equipped HRAs to support driver well-being and enhance overall road safety. The impact of work schedules and the availability of suitable resting areas should be taken into account in interventions aimed at reducing driver weariness, as examined in [9].

Despite the critical role of RAs in promoting road safety, there is a notable deficiency in research focusing on their utilization patterns and the factors influencing their use. While numerous studies have addressed the impact of driver fatigue on accident rates, few have examined how the availability, location, and amenities of RAs affect drivers' decisions to utilize these facilities. However, comprehensive analyses of HRA utilization patterns remain scarce. Understanding these patterns is essential for developing targeted interventions aimed at enhancing RA usage and, consequently, reducing fatigue-related accidents.

Given the growing complexity of transport systems and the increasing availability of infrastructure-related data, data-driven approaches have gained traction as effective tools for understanding and optimizing HRAs usage. Traditional planning models often rely on aggregated traffic volumes and static facility inventories, offering limited insight into temporal and behavioral dynamics. In contrast, modern data analytics enable researchers to capture detailed usage patterns, identify peak demand periods, and evaluate the functional adequacy of existing RA infrastructure. As highlighted by Cai et al. [10] the integration of electronic toll collection (ETC) data and facility-level records can facilitate the estimation of dwell times and occupancy trends in real-world conditions. Such approaches provide the foundation for evidence-based infrastructure planning, especially in regions with heterogeneous traffic flows or uneven facility distribution.

Machine learning methods are increasingly being applied in transportation research to address these analytical needs. Their ability to handle nonlinear relationships and high-dimensional data makes them well-suited for predicting HRA occupancy, identifying underutilized or overburdened facilities, and supporting strategic planning decisions. Previous studies have demonstrated the utility of classifiers such as decision trees, random forests, and boosting algorithms in modeling spatiotemporal traffic dynamics and facility use [8 - 9]. These techniques not only improve prediction accuracy but also offer insights into the relative importance of various infrastructural and contextual factors affecting occupancy levels.

Given the increasing availability of real-time and infrastructure-related data, data-driven approaches have become essential for analyzing HRAs utilization and improving facility planning. Traditional models rarely capture temporal variation or behavioral aspects of HRAs usage. In contrast, modern predictive techniques, such as machine learning, allow for short-term forecasting of parking availability and occupancy states. For instance, Provoost et al. demonstrated that real-time vehicle detection data, when combined with gradient boosting models, significantly improves the prediction of HRAs states, providing valuable input for dynamic traffic and parking management

systems [13]. Similarly, Koesdwiady et al. integrated weather information into deep learning models for traffic flow prediction, showing that external contextual factors can meaningfully enhance predictive accuracy in transportation datasets [14]. These methodological advances are highly relevant to HRAs modeling, where occupancy patterns depend on a complex interplay of infrastructure features, temporal demand, and driver behavior. As interest in intelligent transport systems continues to grow, predictive modeling of HRAs utilization offers a promising direction for improving road safety, optimizing investments, and enhancing user experience.

1.2. Research Gap

Although HRAs have been widely acknowledged as critical infrastructure for mitigating driver fatigue and enhancing road safety, their actual utilization patterns remain poorly understood. Prior studies have predominantly focused on macro-level metrics such as spatial distribution or total facility counts, often neglecting the temporal, behavioral, and functional aspects of RA use. Exploration of the spatial distribution of the service and RAs on the toll motorways in the European Union (EU) presented in [15] revealed a great variability due to different distribution policies. Detailed analysis of the infrastructure and facilities offered in HRAs and the distances between consecutive HRAs was described in many studies on the example of Spain [16], Poland [17] or Lithuania [18]. To determine the quality standards for RA selection, these studies concentrated on analyzing the current state of HRAs in each country. Also, users' preferences for planning parameters of HRAs were introduced in [19].

However, choosing a specific place to rest is not always possible due to the overcrowding of parking lots for truck drivers. In [20] classification method was developed to categorize various types of authorized and unauthorized parking for heavy commercial vehicles (HCV) facing potential parking shortages to accommodate hours of service (HOS) rest requirements. The lack of enough parking spaces during the busiest periods in car parks results in illegal and dangerous parking on entrances and exits and in other unauthorized places, which may pose a risk to road safety. To address parking shortages, researchers have developed various models to predict parking utilization. The models can be based on econometric models when utilizing GPS data to forecast parking demand and identify utilization patterns [21], [22]. Graph neural networks are used to employ spatio-temporal data to predict occupancy rates across multiple sites, enhancing the accuracy of forecasts [23]. Queuing theory models apply statistical methods to model arrival and departure rates, providing real-time probabilistic forecasts of parking occupancy [23].

In the scientific literature, no research was found concerning the RA selection depending on its occupancy in connection with the attractiveness of facilities of these places. Even where HRAs are integrated into broader traffic models, their role is typically limited to static parameters or assumed usage rates, rather than empirically derived occupancy dynamics.

Furthermore, while the application of machine learning in transportation has expanded rapidly in recent years, few studies have employed these methods to model HRAs truck parking occupancy using infrastructure-level attributes. Existing work tends to concentrate on highway congestion prediction or traffic flow forecasting, with limited attention given to facility-specific demand modeling. To the best of our knowledge, no prior research has systematically evaluated and compared supervised classification algorithms for predicting HRAs truck parking occupancy on an hourly basis using a comprehensive dataset combining infrastructural, administrative, and locational features.

This gap is particularly relevant given the increasing availability of high-resolution data from digital monitoring systems and the growing pressure on infrastructure managers to allocate resources efficiently. Addressing this gap through data-driven, interpretable models offers both theoretical and practical contributions to the field of transport infrastructure planning.

1.3. Research Objectives

This study aims to develop and evaluate predictive models using machine learning of HRAs truck parking occupancy using a dataset collected from a digital monitoring system covering selected motorway facilities in Poland. The research also includes an overview of the multi-faceted importance of rest areas on motorways, supported by empirical research and international standards. On this basis it will be possible to determine the extent to which infrastructural, administrative, and locational features can be used to predict hourly occupancy levels with sufficient accuracy and interpretability.

Five supervised machine learning algorithms—Decision Tree, Random Forest, Gradient Boosting, Logistic Regression, and XGBoost—are implemented and compared in the research. The models are evaluated based on their predictive performance as well as their ability to highlight key variables affecting occupancy patterns. Particular emphasis is placed on model explainability, with the goal of identifying which features (e.g., number of parking spaces, proximity to urban areas, presence of services) most strongly influence utilization levels.

By achieving these objectives, the study seeks to contribute to the development of data-driven approaches for transport infrastructure management. The findings are expected to support more informed decision-making in the design, upgrading, and spatial allocation of HRAs, ultimately contributing to safer and more efficient road transport systems.

1.4. Manuscript Structure

The remainder of this manuscript is organized as follows: Section 1 presents the introduction, outlining the background and motivation for the study, followed by a review of relevant literature. It identifies existing research gaps and defines the main objectives of the study, concluding with a summary of the manuscript structure. Section 2 describes the materials and methods employed in the research. It includes a detailed description of the study area and elaborates on the research methodology, including data collection and machine learning techniques used for parking occupancy prediction. Section 3 presents the results of the study, showcasing the performance of the predictive models and the key findings related to truck parking occupancy patterns. Section 4 provides a discussion of the results, placing them in the context of previous research, interpreting their implications, and highlighting both the strengths and limitations of the approach. Section 5 concludes the manuscript with a summary of the main findings, contributions of the research, and suggestions for future work. This structured approach ensures a clear and logical flow of information from problem definition to practical conclusions.

2. Materials and Methods

2.1. Research Area

The geographical area of the study covered the southern part of Poland (Figure 1). The analysis area is marked with a red circle, however, not all points visible as numbers in the blue circled locations were taken into account during the research. The analysis covered 69 HRAs located in 39 towns mainly along the A4 motorway section with access expressways in four voivodeships: Lower Silesia (22), Lesser Poland (12), Opole Voivodeship (6) and Silesia (29). The points analyzed included HRAs of classes I, II and III managed by the Polish General Directorate for National Roads and Motorways, as well as independent parking lots of companies that provide parking spaces for trucks.

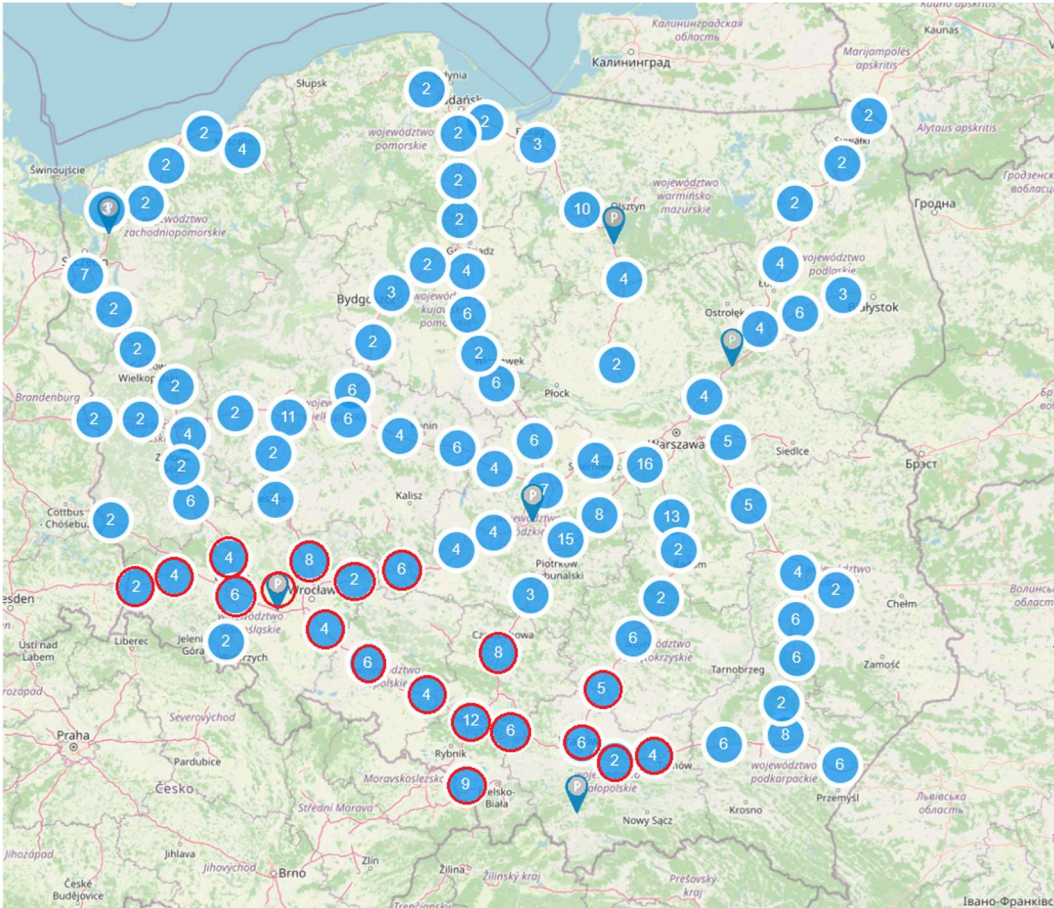


Figure 1. Highway rest areas in Poland with geographical area of analysis in southern part (red circles). Source: own research based on [24].

There are over 410 HRAs in Poland, of which the General Directorate for National Roads and Motorways manages 157 points with a basic function and the remaining 112 areas with a commercial function. Polish HRAs are divided into classes of different sizes which is adapted to the average daily traffic in a year. When designing rest stations, they should be built alternately to guarantee passengers the best functionality, taking into account the facilities for each class, presented in Table 1.

Table 1. Characteristics of facilities for individual classes of HRAs (+ exists, - does not exist).

Type of facility	HRA class I	HRA class II	HRA class II
Area surface	up to 1.5 ha	1.5-3.0 ha	3.0-4,5 ha
Parking spaces for passenger cars	+	+	+
Parking spaces for trucks	+	+	+
Sanitary facilities	+	+	+
Water collection point	+	+	+
Pedestrian bridge for car inspection	+	+	+
Fuel station	-	+	+
Catering services (fast - food)	-	+	+
Service station	-	-	+
Catering services (restaurant)	-	-	+
Accommodation services	-	-	+

Theoretical distances are specified that should separate specific categories of HRAs. They should be a minimum of 15 km between adjacent HRAs, between HRA type II and III from 50 to 75 km, and

HRA of the same type on the same side of the motorway should be separated by 100-150 km [25], as shown schematically in the Figure 2.

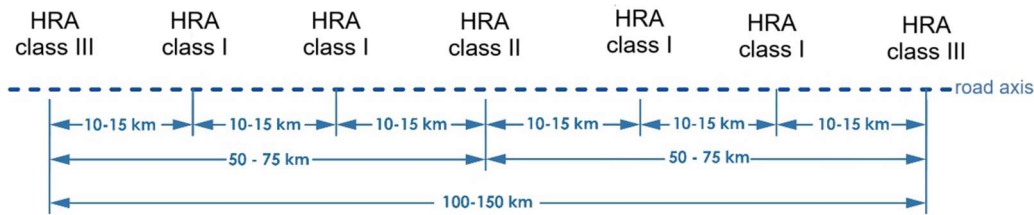


Figure 2. Recommended distances between individual categories of HRAs. Source: own research based on [25].

For the HRA system to be as functional as possible, the high speed at which drivers travel on the motorway must also be taken into account. It is therefore necessary to enable them to safely brake the vehicle and smoothly join the traffic when returning to the motorway. This is possible thanks to the use of an extended communication system. This type of system also allows drivers traveling on the motorway to notice the place more easily. The recommended scheme for shaping the HRA surface, taking into account main functions of the facilities, is shown in Figure 3.

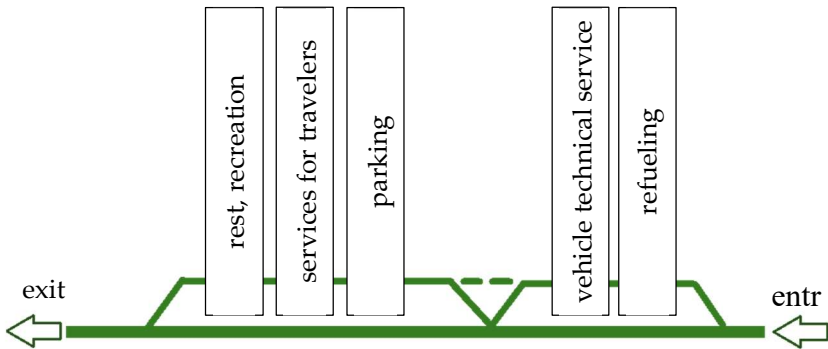


Figure 3. Recommended scheme for shaping the communication system and the arrangement of facilities included in the HRA. Source: own research based on [25].

Spatial development plan of an exemplary polish HRA (Żarska Wieś) including the layout of individual facilities is shown in Figure 4.

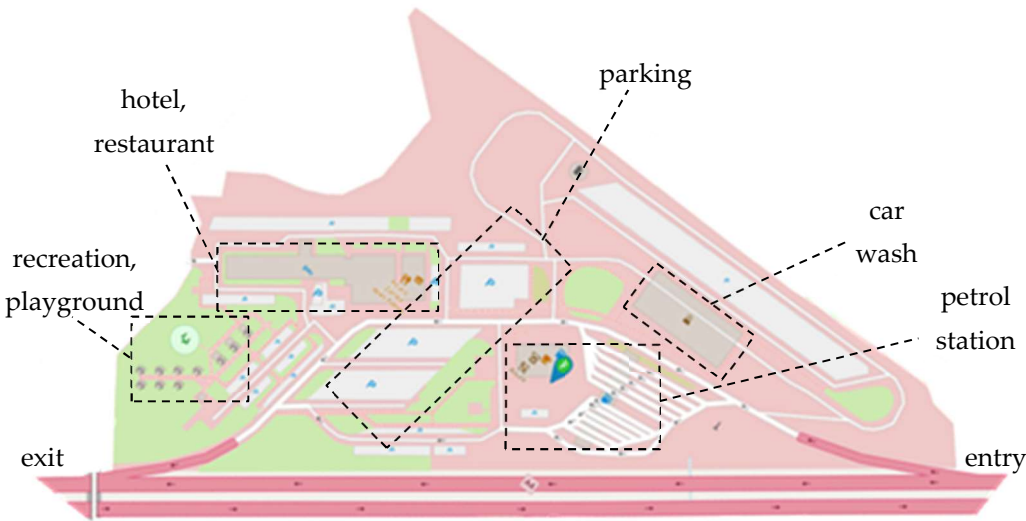


Figure 4. Spatial development plan of an exemplary HRA (Żarska Wieś) including the layout of individual facilities. Source: own research based on [24].

When analyzing the future route of the road and the location of HRAs by category in Poland, not only the requirements resulting from the regulations are taken into account, but also terrain conditions, traffic forecasts, the location of other places on the network or the distance from agglomerations or state borders. The location of HRAs together with the proposed route of the road are subject to public consultations.

2.2. Research Methodology

The first stage of the research was the analysis of 69 HRAs selected in the previous stage in terms of their type and services offered. Similarly, to the General Directorate for National Roads and Motorways, an analysis of existing features describing infrastructural, administrative, and locational characteristics was made for each HRA. It included: name, branch, voivodeship, manager type, location, category, geographical coordinates of road technical class and number, number of parking lots for cars, trucks and buses, toilets, fuel station, food service point, accommodation, electric vehicle charging stations, hydrogen, CNG and LNG refueling stations, security, fence, video monitoring, lighting, wash, vehicle repair workshop, liquid waste discharge points, shower, and snow removal ramps.

In this study, data originating from a dedicated application that monitors highway rest area occupancy were used to create a single, standardized dataset. The selection of an application for tracking parking occupancy was preceded by interviews with truck drivers and a ranking of similar mobile applications in GooglePlay. The LKW.APP mobile application was designed in 2022 to facilitate the work of drivers, to reduce their stress level, but also to reduce the cost of transport when planning a stopover and searching for a free space at a truck parking lot. The app allows to check the availability of truck parking lots across 40 thousand European parking lots for free, using color coding, allowing to plan the route in real time and up to 15 hours in advance. In addition, it allows you to find a parking lot for trucks with special services, adapting the route to specific vehicle parameters (e.g. avoiding narrow roads). The LKW.APP application, available in 16 languages, is quite popular, with over 100 thousand downloads, users in 44 countries and a rating of 4.8 on GooglePlay. Screenshots of the three most important features of the app are shown in Figure 5.

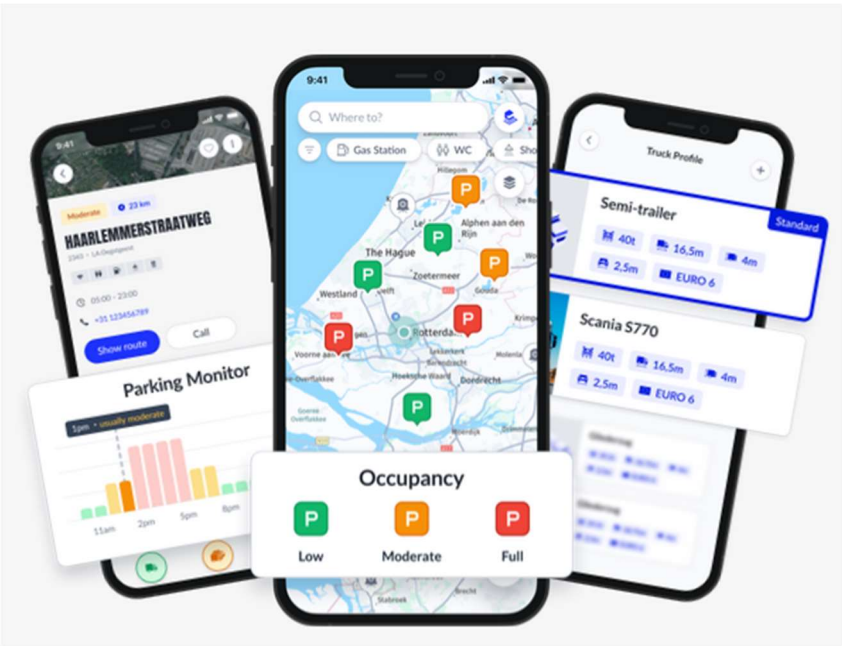


Figure 5. LKW.APP mobile application with three key functions: parking monitor, parking occupancy and truck profile. Source: [21].

The dataset of 10,740 observations for the research comprised multiple worksheets, each capturing HRAs truck parking hourly occupancy information for random days of the week of 69 rest areas at the beginning of the year 2025. After discarding any worksheets designated as templates, columns encoding the hourly measurements were distinguished from those containing metadata. The resulting data were transformed from a wide format—where each rest area was represented by multiple hour-specific columns—into a long format, wherein each row corresponded to a distinct rest area–hour pair. Invalid or missing entries were removed, and all columns were merged so that no features were lost, with unavailable fields marked as empty. Additionally, the original occupancy markers “L”, “M”, and “F” were standardized to three English descriptors—“low,” “moderate,” and “full”—mapping “L” to “low,” “M” to “moderate,” and “F” to “full”.

The dataset was further examined using basic descriptive statistics to provide an initial overview of variable distributions and frequencies. This was achieved with the `describe()` function from the pandas library, which provides summary measures tailored to the type of data being analyzed [26]. For numerical variables, the output includes the count of non-null values, mean, standard deviation, minimum, maximum, and quartile values (25%, 50%, and 75%), allowing for the identification of central tendency and variability [27]. In the case of categorical variables, the summary returns the number of valid entries, the number of unique categories, the most frequent category (mode), and its frequency, which is useful for detecting dominant labels and structural imbalance [28]. These descriptive outputs serve as a foundation for exploratory data analysis and guide further statistical modeling.

Python is increasingly regarded as a leading programming language for data analysis and scientific computing due to its readability, versatility, and wide array of specialized libraries [29]. In this study, Python was used as the primary platform for the entire analytical workflow, encompassing data cleaning, transformation, visualization, and predictive modeling. Key packages included pandas [30] for data manipulation, matplotlib [31] for visual representation, and scikit-learn [32] for implementing standard machine learning models. Additionally, xgboost [33] was employed for building high-performance gradient-boosted classifiers. The modularity of the Python environment and the interactivity of Jupyter Notebooks [34] enabled transparent and reproducible experimentation, making it especially suitable for research applications involving transportation data [35].

To predict HRAs truck parking occupancy levels, five supervised classification algorithms were implemented using Python's machine learning ecosystem. The models included: Decision Tree, Random Forest, Gradient Boosting, Logistic Regression, and XGBoost classifiers. Decision Trees serve as a fundamental approach that recursively split data based on feature thresholds to minimize impurity in child nodes [36]. Random Forest expands on this concept by aggregating the predictions of multiple randomized trees to improve generalization and reduce overfitting [37]. Gradient Boosting constructs an ensemble in a sequential manner, where each tree is trained to correct the residual errors of its predecessor, often resulting in high predictive performance [38]. Logistic Regression, despite its simplicity, is a robust linear model commonly used in classification tasks [39]. Lastly, the XGBoost model is an advanced boosting technique known for its computational efficiency and regularization features, making it particularly effective for tabular data. All classifiers were evaluated using a consistent train-test split, with the same preprocessed dataset and encoded categorical variables. The performance of each model was assessed through standard classification metrics, including accuracy, precision, recall, and F1-score.

To assess the predictive performance of the trained models, a set of widely accepted evaluation metrics for multi-class classification tasks was employed. These included accuracy, precision, recall, and F1-score, all computed using functions available in the scikit-learn environment. Accuracy provided a general measure of correct predictions across all classes, while precision and recall offered a more detailed view of model performance for each occupancy category—particularly important due to class imbalance. The F1-score, being the harmonic mean of precision and recall, served as a balanced metric accounting for both false positives and false negatives, and is especially suitable

when the cost of misclassification differs across classes. Such metrics are standard practice in model evaluation for classification problems in transportation research and beyond [40]. For transparency, the results were reported separately for each class (low, medium, high), along with macro and weighted averages. Finally, models were ranked by their accuracy scores and visualized using a comparative bar chart to facilitate direct interpretation and model selection.

To assess model performance, a comprehensive set of evaluation metrics was applied, including accuracy, precision, recall, and F1-score—computed separately for each occupancy class using the `classification_report()` function from the scikit-learn package. These measures are commonly used in multi-class classification tasks and offer detailed insight into model behavior. Additionally, feature importance analysis was conducted to identify the most influential variables. This was based on the `feature_importances_` attribute of a trained XGBoost model, which reflects each feature's contribution to error reduction across the boosting process. Only features with importance values above 0.001 were retained and visualized using a horizontal bar chart for better readability. Such methods support interpretability and transparency in applied machine learning, particularly when decisions impact infrastructure and policy planning [41]. The full analytical pipeline, including data preprocessing and model implementation in Python, is available in a public repository: <https://github.com/BudzynskiA/rest-area-occupancy-ml>.

3. Results

To investigate the usage of HRAs, a dataset of 10,740 observations and 33 variables was prepared. It included both numerical and categorical features, such as the number of parking spaces for passenger vehicles, buses, and trucks, as well as administrative and locational attributes like region, administrator, municipality, direction, and road classification. Additionally, the dataset covered the availability of various facilities and services, including restaurants or bistros, accommodation, toilets, showers, video surveillance, lighting, fencing, guarding, electric vehicle charging stations, hydrogen, CNG, and LNG refueling infrastructure, fuel stations, and workshops. Each row represented a unique rest area–hour combination, with occupancy classified into one of three standardized categories: low, medium, or high.

Error! Reference source not found. presents the overall distribution of truck parking lots occupancy levels recorded across all observations. Medium occupancy emerges as the most frequent category, with 5,455 observations, followed by low occupancy (4,002 cases), and high occupancy, which appears in only 1,283 instances. This distribution suggests that moderate usage predominates among highway rest areas, while instances of high congestion are relatively rare.

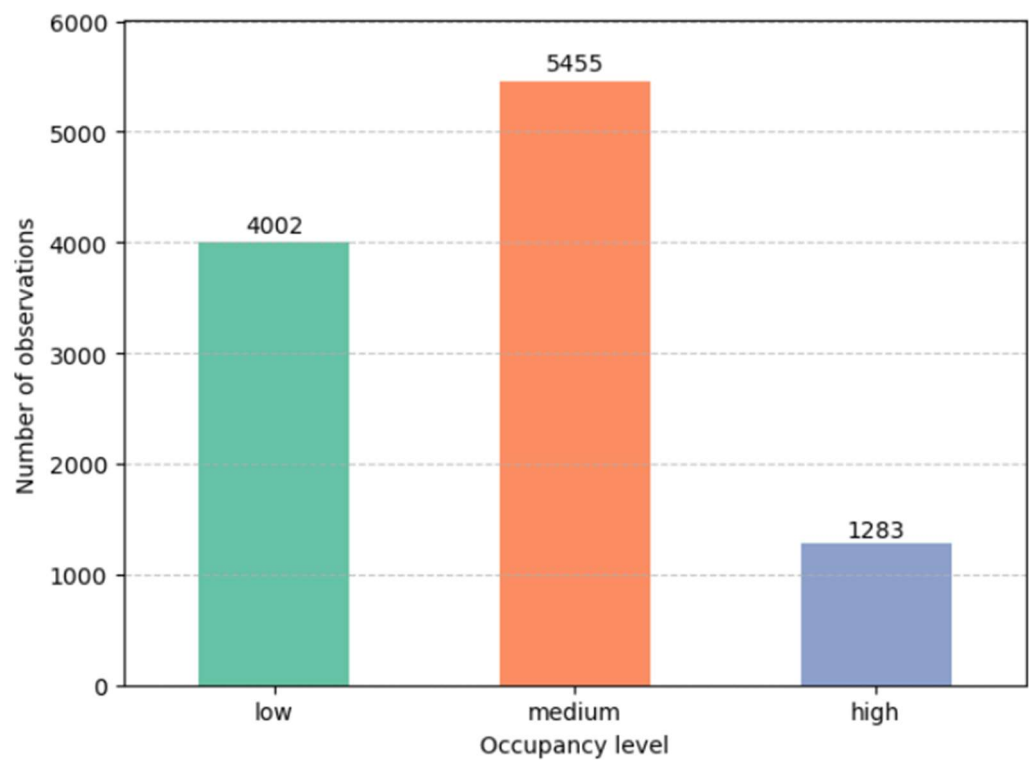


Figure 6. Overall occupancy level distribution of HRAs.

Error! Reference source not found. presents the hourly distribution of rest area occupancy across all observations. A distinct temporal pattern emerges, with low occupancy levels dominating the early part of the day, peaking between 10:00 and 12:00, followed by a gradual decline during the afternoon and evening hours. Medium occupancy remains relatively stable throughout the day, with a slight upward trend from morning to evening. In contrast, high occupancy levels are consistently low during the morning but increase sharply after 17:00, reaching their highest values around 21:00.

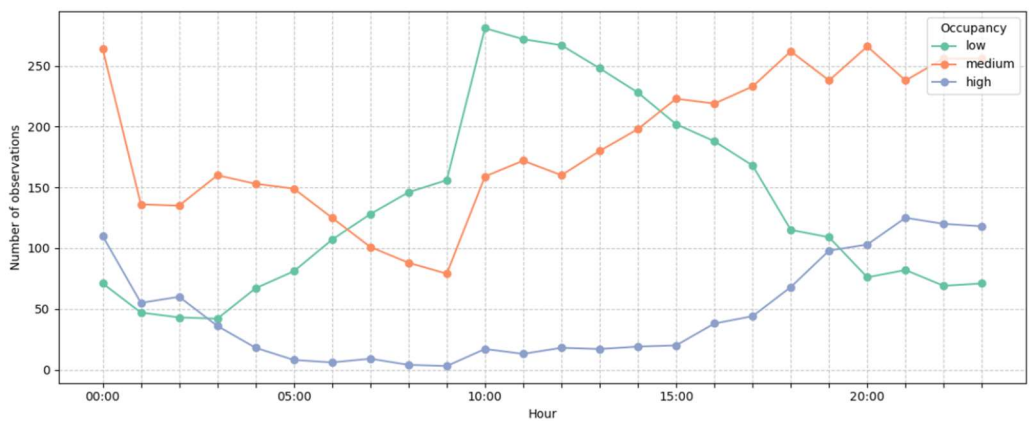


Figure 7. Hourly occupancy trends.

Error! Reference source not found. presents the percentage distribution of occupancy levels across four southern regions of Poland. Medium occupancy is the dominant category in all voivodeships, ranging from 44.0% in Lesser Poland to 66.7% in Opole. Low occupancy is most prevalent in Lesser Poland (48.1%), while Silesia shows the highest proportion of high occupancy (14.4%).

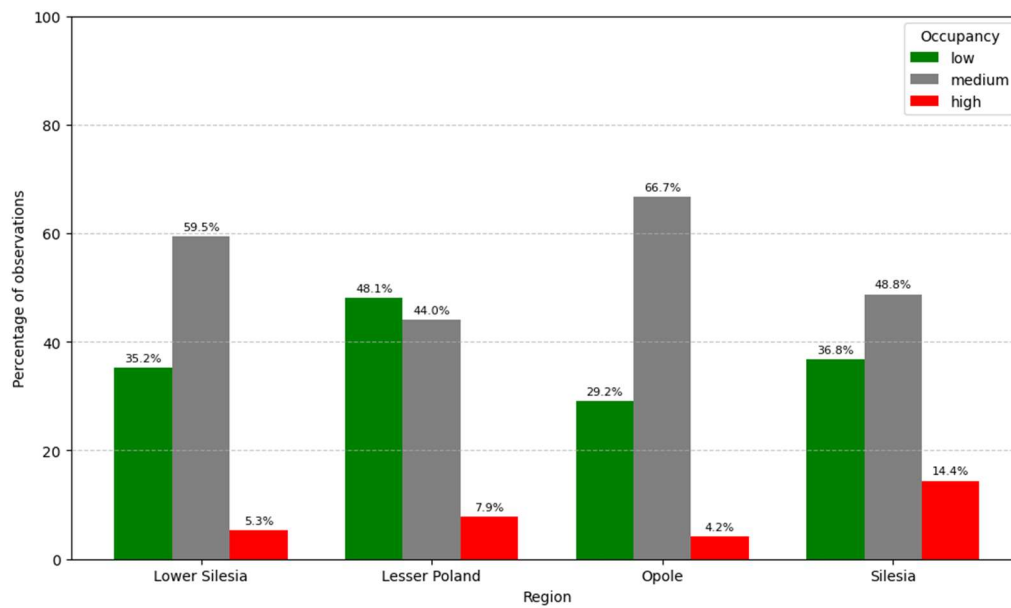


Figure 8. Percentage occupancy by region.

Geographic coordinates in all voivodeships of HRAs in relation to their total parking lots is shown in Figure 9.

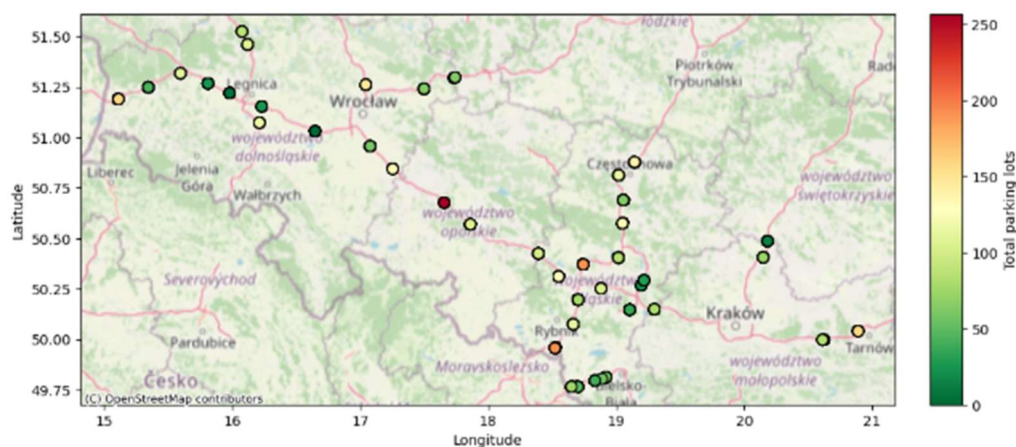


Figure 9. Map showing the geographic coordinates of HRAs and the number of associated parking lots.

Figure 10 illustrates the occupancy levels in relation to the presence of specific facilities at highway rest areas. The presence of certain facilities, such as fuel stations, restaurants, video monitoring, and lighting, correlates with a higher proportion of medium and high occupancy levels. Rest areas equipped with fuel stations and restaurants show elevated frequencies of high occupancy.

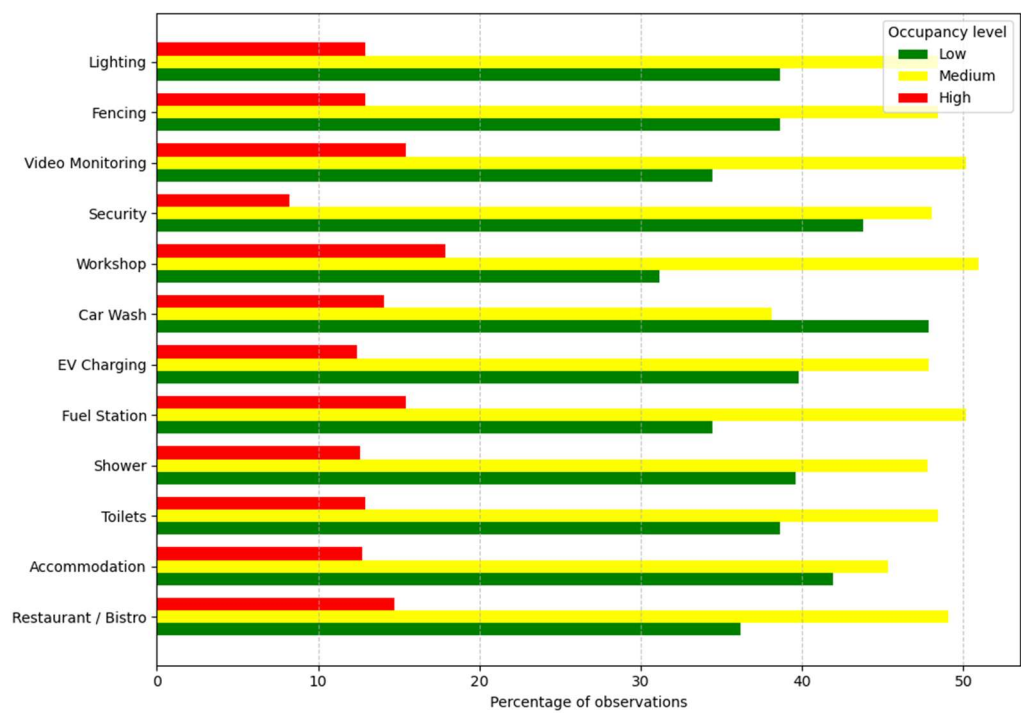


Figure 10. Percentage distribution of occupancy levels by facility.

To further explore how specific amenities may influence rest area usage, occupancy levels were encoded numerically (1 = low, 2 = medium, 3 = high), and the average was calculated for facilities with and without selected features. As shown in **Error! Reference source not found.**, some elements, such as workshop, video monitoring, and fuel station, are associated with higher average occupancy, suggesting that the presence of these services might enhance the utility or attractiveness of the site. Conversely, sites with car wash or security reported slightly lower average occupancy, which may reflect either operational characteristics or user preferences.

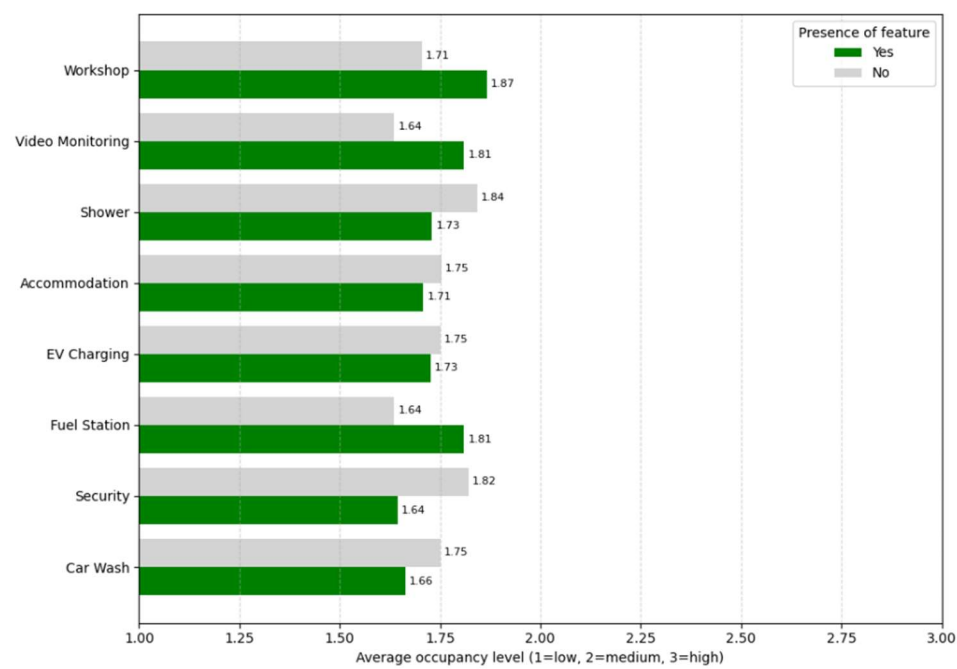


Figure 11. Average occupancy level for rest areas with and without each feature.

Error! Reference source not found. presents the accuracy scores of the five classification models used to predict truck parking occupancy levels. The Gradient Boosting classifier achieved the highest accuracy (0.66), indicating its relatively strong performance on this dataset. XGBoost and Random Forest followed with accuracies of 0.59 and 0.58, respectively. Logistic Regression and Decision Tree models obtained lower accuracy values, at 0.57 and 0.56. All models were trained and evaluated using the same preprocessed dataset and classification framework, allowing for a direct comparison of predictive performance.

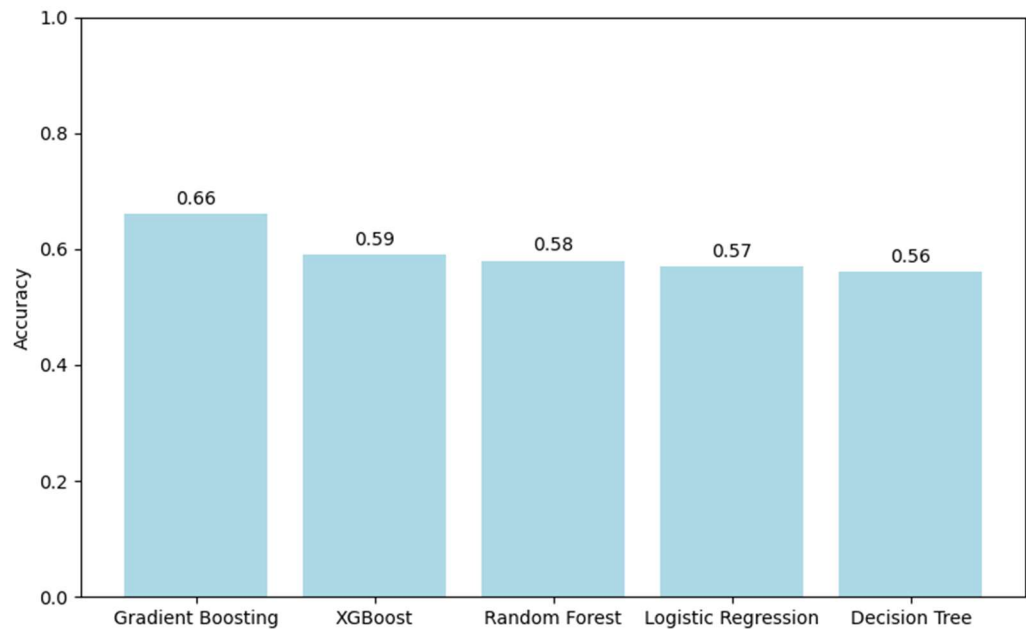


Figure 12. Model accuracy comparison.

Error! Reference source not found. presents the feature importance rankings derived from the Gradient Boosting model used to predict rest area occupancy levels.

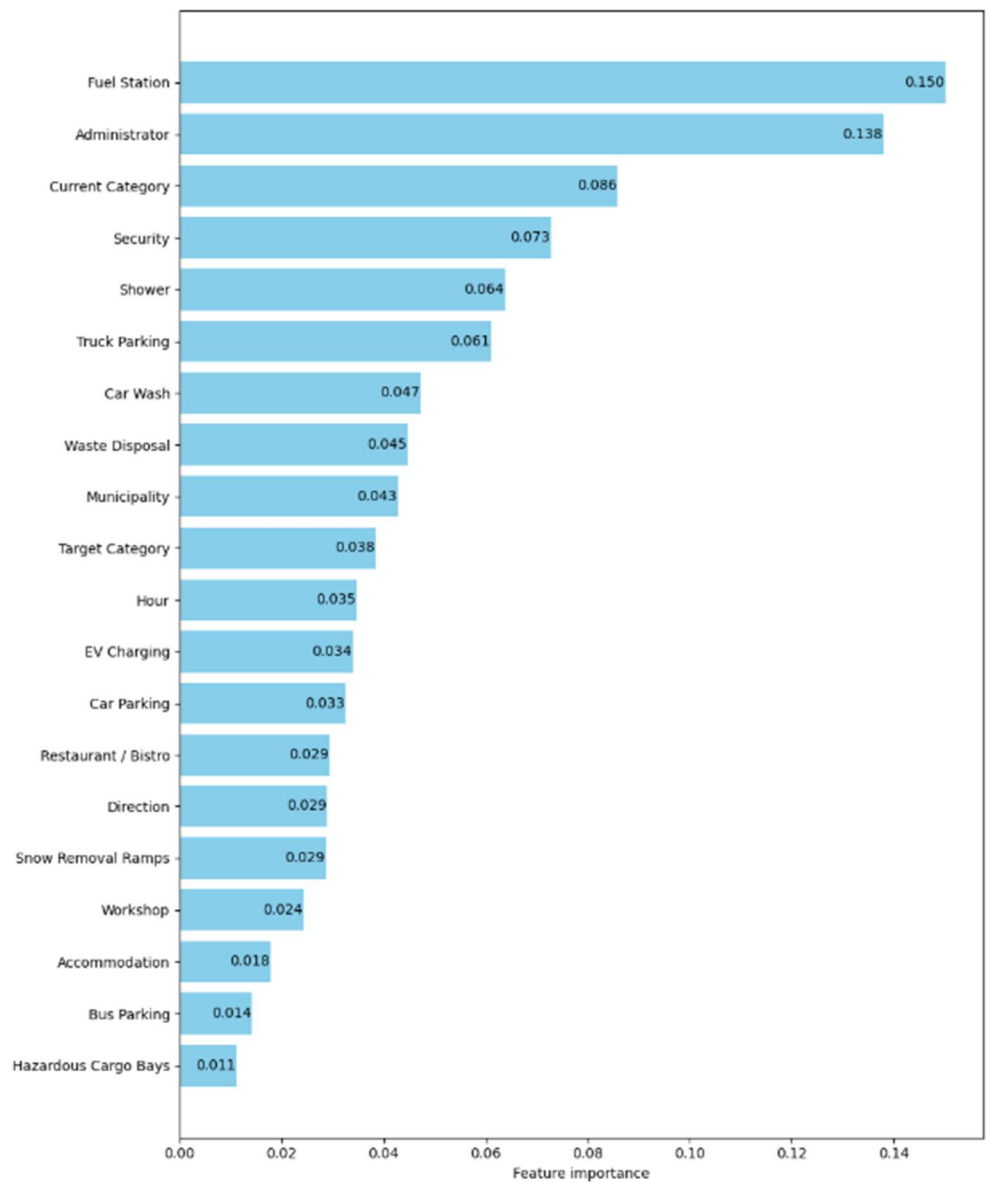


Figure 13. Feature importance.

The analysis indicates that the presence of a fuel station is the most significant predictor, followed by factors such as administrator identity, current rest area category, security, showers, truck parking, and waste disposal facilities. These findings suggest that the availability of comprehensive amenities and effective management play a crucial role in attracting users to rest areas. This observation is consistent with the literature, which emphasizes that well-equipped rest areas significantly enhance traveler comfort and safety, thereby increasing their utilization.

4. Discussion

Current standards for the implementation of infrastructure investments in the scope of HRAs assume that the basic infrastructure of each of them should consist of a parking lot for passenger cars, trucks and buses, toilets, places to rest, a playground for children and an outdoor gym. The development of this infrastructure in Poland aims to improve the comfort and safety of travelers, especially professional drivers, and to provide them with the necessary services during long journeys. Work is underway to modernize (especially in terms of increasing the number of snow removal points) and expand existing HRAs, and the construction of new ones is planned, in line with national and European requirements, including electric vehicle charging stations.

The results of this study confirm that machine learning models can be effectively applied to predict hourly occupancy levels at highway rest areas using structured, infrastructure-related data. Among the tested algorithms, the best-performing model demonstrated strong predictive capabilities, particularly in capturing complex interactions between facility characteristics and usage dynamics. Variables such as total parking capacity, proximity to urban centers, and the availability of essential services consistently emerged as the most influential features, suggesting that both spatial and functional factors contribute significantly to rest area demand.

Moreover, the observed predominance of medium occupancy levels and the relatively low occurrence of full capacity underscore temporal occupancy patterns and their implications for facility management, aligning with previous research indicating that rest area utilization rates typically range between 8.4% and 12.3% of mainline traffic, with peak usage around midday [42]. These findings suggest that most highway rest areas operate below maximum capacity, experiencing higher congestion only during specific time windows. This temporal pattern underscores the importance of effective facility management to address peak periods without incurring unnecessary expansion costs.

The influence of specific amenities on occupancy levels also supports findings from previous studies suggesting that enhanced rest area infrastructure significantly shapes driver preferences and usage patterns [43]. In particular, facilities offering fuel, food, or improved safety features tend to attract more users, indicating that investment in comprehensive services can yield measurable increases in rest area utilization. These insights are crucial for infrastructure planning strategies aiming to optimize not just capacity, but also functional attractiveness.

The regional variation in occupancy levels observed across voivodeships highlights critical implications for infrastructure planning and capacity management, particularly in Silesia and Lesser Poland. Lesser Poland exhibited relatively high proportions of low occupancy, potentially indicating underutilization of available infrastructure, whereas Silesia showed elevated levels of high occupancy, suggesting localized capacity constraints. These disparities align with prior analyses emphasizing the importance of spatial dynamics in highway traffic management and infrastructure planning [44].

The diurnal occupancy cycle identified in this study, characterized by low usage in the morning and increased congestion in the evening, aligns with previous findings on rest area stop behavior [45]. Recent research also underscores the importance of incorporating temporal dynamics into predictive models of rest area usage, suggesting that hybrid approaches can effectively capture these cyclical patterns [46]. Studies have shown that short-term, low-density stops typically prevail in the morning, while congestion-related or long-duration stops are more common in the evening hours. These temporal dynamics emphasize the need to integrate time-of-day variations into predictive models of rest area usage, as hybrid approaches have demonstrated effectiveness in capturing such cyclical patterns. This pattern suggests that time-of-day dynamics should be integrated into the design and management of rest area facilities to address varying levels of demand effectively.

Furthermore, the comparative evaluation of five machine learning algorithms introduces a clearer transition to the comparison, emphasizing the role of Gradient Boosting as the most effective model for capturing non-linear interactions in the dataset reaffirmed the practical advantages of ensemble methods for infrastructure-related classification problems. The Gradient Boosting model outperformed other classifiers, achieving the highest accuracy and effectively capturing non-linear interactions among infrastructure features. This outcome aligns with existing studies demonstrating the robustness of boosting algorithms in transportation-related prediction tasks [11]. Random Forest and XGBoost also yielded competitive results, while Logistic Regression and Decision Tree models performed less effectively, consistent with prior reports highlighting their limited effectiveness in capturing nuanced behavioral patterns in transportation datasets.

The findings also emphasize the critical role of feature engineering and model interpretability within the predictive framework employed in this study, and model interpretability in infrastructure forecasting. Recent meta-analyses emphasize the value of combining regression, decision trees, and

deep learning methods to manage the heterogeneous nature of rest area data [47]. Although the present study focused on interpretable classification algorithms, future research could explore hybrid models that integrate temporal learning architectures to capture more complex spatial and temporal dynamics.

Despite these promising outcomes, certain limitations should be acknowledged, particularly regarding data coverage, model generalizability, and the integration of external factors such as weather, seasonal variation, and real-time traffic conditions. The dataset was geographically limited to southern Poland and did not account for external factors such as weather or seasonal variation, potentially affecting model generalizability. Additionally, the occupancy data, while consistent, were derived from digital monitoring systems and may not fully capture nuanced behavioral patterns. Finally, while the feature importance analysis provides a degree of interpretability, further research is necessary to enhance transparency and build stakeholder trust in predictive tools employed for public infrastructure planning.

5. Conclusions

As traffic volumes increase and long-distance travel becomes more frequent, rest areas serve as crucial nodes for maintaining traffic safety, enhancing driver health and balance, and supporting logistical and environmental goals. They are strategic components of modern road infrastructure. Equipping them appropriately to user needs and evaluating predictive models of rest area occupancy is crucial for the development of sustainable and utilitarian transport networks based on SSTPAs programme.

This study demonstrated the applicability of supervised machine learning models for predicting hourly occupancy levels at highway rest areas using structured data on facility characteristics and location. Among the tested algorithms, the best-performing model achieved high predictive accuracy and identified several key features—such as parking capacity, proximity to urban areas, and availability of services—as major determinants of rest area utilization.

By integrating infrastructure-specific attributes into a classification framework, the study moves beyond traditional planning approaches that rely on static assumptions or aggregate demand estimates. The results provide a replicable methodology for transport agencies seeking to monitor, evaluate, or redesign rest area networks based on empirical usage patterns.

In addition to confirming the relevance of machine learning in infrastructure analytics, the study highlights the need for further research incorporating real-time data sources and external factors such as weather or seasonality. Expanding the geographic scope and testing model generalizability in other contexts would enhance the robustness and transferability of the findings.

Overall, the research contributes to a growing body of evidence supporting data-driven transport infrastructure planning and offers actionable insights for improving the efficiency, safety, and responsiveness of rest area provision.

Author Contributions: Conceptualization, A.B. and M.C.; methodology, A.B. and M.C.; formal analysis, A.B. and M.C.; investigation, A.B. and M.C.; data curation, M.C.; writing—original draft preparation, A.B. and M.C.; writing—review and editing, A.B. and M.C.; visualization, A.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data are contained within the article. The full analytical pipeline, including data preprocessing and model implementation in Python, is available in a public repository: <https://github.com/BudzynskiA/rest-area-occupancy-ml>.

Acknowledgments: The authors would like to thank the reviewers for their valuable and perceptive comments, which have enhanced the paper's quality and will assist the authors in advancing their study in this area.

Conflicts of Interest: The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

ETC	electronic toll collection
HCV	Heavy Commercial Vehicles
HoS	Hours of Service
HRAs	Highway Rest Areas
RAs	Rest Areas
SSTPAs	Safe and Secure Truck Parking Areas

References

1. G. Cheng, R. Cheng, Y. Pei, and J. Han, ‘Research on Highway Roadside Safety’, *J. Adv. Transp.*, vol. 2021, pp. 1–19, Jun. 2021. <https://doi.org/10.1155/2021/6622360>.
2. EU, *Commission Delegated Regulation (EU) 2022/1012 of 7 April 2022 supplementing Regulation (EC) No 561/2006 of the European Parliament and of the Council with regard to the establishment of standards detailing the level of service and security of safe and secure parking areas and to the procedures for their certification*. 2022.
3. World Health Organization, *Global status report on road safety 2018*. Geneva: World Health Organization, 2018. Accessed: Apr. 14, 2025. [Online]. Available: <https://iris.who.int/handle/10665/276462>
4. I. D. Brown, ‘Driver Fatigue’, *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 36, no. 2, pp. 298–314, Jun. 1994. <https://doi.org/10.1177/001872089403600210>.
5. A. McArthur, J. Kay, P. T. Savolainen, and T. J. Gates, ‘Effects of Public Rest Areas on Fatigue-Related Crashes’, *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2386, no. 1, pp. 16–25, Jan. 2013. <https://doi.org/10.3141/2386-03>.
6. T. L. Bunn, S. Slavova, and P. J. Rock, ‘Association between commercial vehicle driver at-fault crashes involving sleepiness/fatigue and proximity to rest areas and truck stops’, *Accid. Anal. Prev.*, vol. 126, pp. 3–9, May 2019. <https://doi.org/10.1016/j.aap.2017.11.022>.
7. F. Alkhatni, S. Z. Ishak, and A. Milad, ‘Characteristics and Potential Impacts of Rest Areas Proximate to Roadways: A Review’, *Open Transp. J.*, vol. 15, no. 1, pp. 260–271, Dec. 2021. <https://doi.org/10.2174/1874447802115010260>.
8. A. M. Crizzle, R. Toxopeus, and J. Malkin, ‘Impact of limited rest areas on truck driver crashes in Saskatchewan: a mixed-methods approach’, *BMC Public Health*, vol. 20, no. 1, p. 971, Dec. 2020. <https://doi.org/10.1186/s12889-020-09120-7>.
9. X. Ren, E. Pritchard, C. Van Vreden, S. Newnam, R. Iles, and T. Xia, ‘Factors Associated with Fatigued Driving among Australian Truck Drivers: A Cross-Sectional Study’, *Int. J. Environ. Res. Public Health*, vol. 20, no. 3, p. 2732, Feb. 2023. <https://doi.org/10.3390/ijerph20032732>.
10. Q. Cai, D. Yi, F. Zou, Z. Zhou, N. Li, and F. Guo, ‘Recognition of Vehicles Entering Expressway Service Areas and Estimation of Dwell Time Using ETC Data’, *Entropy*, vol. 24, no. 9, p. 1208, Aug. 2022. <https://doi.org/10.3390/e24091208>.
11. G. Dai, J. Tang, and W. Luo, ‘Short-term traffic flow prediction: An ensemble machine learning approach’, *Alex. Eng. J.*, vol. 74, pp. 467–480, Jul. 2023. <https://doi.org/10.1016/j.aej.2023.05.015>.
12. Y. Yu, Q. Shang, and T. Xie, ‘A Hybrid Model for Short-Term Traffic Flow Prediction Based on Variational Mode Decomposition, Wavelet Threshold Denoising, and Long Short-Term Memory Neural Network’, *Complexity*, vol. 2021, no. 1, p. 7756299, Jan. 2021. <https://doi.org/10.1155/2021/7756299>.
13. J. Provoost, L. Wismans, S. V. der Drift, A. Kamilaris, and M. V. Keulen, ‘Short Term Prediction of Parking Area states Using Real Time Data and Machine Learning Techniques’, Nov. 29, 2019, *arXiv*: arXiv:1911.13178. <https://doi.org/10.48550/arXiv.1911.13178>.
14. A. Koesdwiady, R. Soua, and F. Karray, ‘Improving Traffic Flow Prediction With Weather Information in Connected Cars: A Deep Learning Approach’, *IEEE Trans. Veh. Technol.*, vol. 65, no. 12, pp. 9508–9517, Dec. 2016. <https://doi.org/10.1109/TVT.2016.2585575>.
15. H. Pérez-Acebo, A. Romo-Martín, and D. J. Findley, ‘Spatial distribution and the facility evaluation of the service and rest areas in the toll motorway network of the European Union’, *Appl. Spat. Anal. Policy*, vol. 15, no. 3, pp. 821–845, Sep. 2022. <https://doi.org/10.1007/s12061-021-09421-3>.

16. A. Romo-Martín and H. Pérez-Acebo, 'Analysis of the Location of Service and Rest Areas and their facilities in Spanish paying motorways', *Transp. Res. Procedia*, vol. 33, pp. 4–11, 2018. <https://doi.org/10.1016/j.trpro.2018.10.069>.
17. H. Pérez-Acebo and A. Romo-Martín, 'Service and rest areas in toll motorways in Poland: study of distribution and facilities', *Transp. Probl.*, vol. 14, no. 2, pp. 155–164, 2019. <https://doi.org/10.20858/tp.2019.14.2.14>.
18. J. Kolodinskaja and L. Bertulienė, 'Layout of Rest Areas and Their Infrastructure Development in the South-Eastern Region of Lithuania', *Balt. J. Road Bridge Eng.*, vol. 15, no. 3, pp. 130–145, Aug. 2020. <https://doi.org/10.7250/bjrbe.2020-15.488>.
19. A. Hami and A. Nojavan, 'Rest Areas Management; the Effect of Demographic Information into Users' Preferences for Planning Parameters of Rest Areas', *Int. J. Archit. Eng. Urban Plan.*, vol. 30, no. 1, Jun. 2020. <https://doi.org/10.22068/ijaup.30.1.97>.
20. E. A. Nevland, K. Gingerich, and P. Y. Park, 'A data-driven systematic approach for identifying and classifying long-haul truck parking locations', *Transp. Policy*, vol. 96, pp. 48–59, Sep. 2020. <https://doi.org/10.1016/j.tranpol.2020.04.003>.
21. K. Haque, S. Mishra, R. Paleti, M. M. Golias, A. A. Sarker, and K. Pujats, 'Truck Parking Utilization Analysis Using GPS Data', *J. Transp. Eng. Part Syst.*, vol. 143, no. 9, p. 04017045, Sep. 2017. <https://doi.org/10.1061/JTEPBS.0000073>.
22. A. Budzyński and M. Cieśla, 'Application of a machine learning model for forecasting freight rate in road transport', *Sci. J. Silesian Univ. Technol. Ser. Transp.*, vol. 126, pp. 23–48, Mar. 2025. <https://doi.org/10.20858/sjsutst.2025.126.2>.
23. R. Tamaru, Y. Cheng, S. Parker, E. Perry, B. Ran, and S. Ahn, 'Truck Parking Usage Prediction with Decomposed Graph Neural Networks', 2024, *arXiv*. <https://doi.org/10.48550/ARXIV.2401.12920>.
24. 'OpenStreetMap Project. Available online: <https://www.openstreetmap.org/> (accessed on 20 April 2025).'
25. J. Biedrońska, Ed., *Projektowanie obiektów motoryzacyjnych*, Wyd. 2. in Monografia / [Politechnika Śląska], no. 262. Gliwice: Wydawnictwo Politechniki Śląskiej, 2010.
26. D. S. Moore, G. P. McCabe, and B. A. Craig, *Introduction to the practice of statistics*, Ninth edition. New York: W.H. Freeman, Macmillan Learning, 2017.
27. A. P. Field, *Discovering statistics using IBM SPSS statistics: and sex and drugs and rock 'n' roll*, 4th edition. Los Angeles: Sage, 2013.
28. W. McKinney, *Python for data analysis: data wrangling with pandas, NumPy, and Jupyter*, Third edition. Beijing: O'Reilly, 2022.
29. J. D. Kelleher and B. Tierney, *Data science*. in The MIT Press essential knowledge series. Cambridge, Massachusetts London, England: The MIT Press, 2018.
30. W. McKinney, 'Data Structures for Statistical Computing in Python', presented at the Python in Science Conference, Austin, Texas, 2010, pp. 56–61. <https://doi.org/10.25080/Majora-92bf1922-00a>.
31. J. D. Hunter, 'Matplotlib: A 2D Graphics Environment', *Comput. Sci. Eng.*, vol. 9, no. 3, pp. 90–95, 2007. <https://doi.org/10.1109/MCSE.2007.55>.
32. F. Pedregosa et al., 'Scikit-learn: Machine Learning in Python', 2011. <https://doi.org/10.48550/ARXIV.1201.0490>.
33. T. Chen and C. Guestrin, 'XGBoost: A Scalable Tree Boosting System', in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco California USA: ACM, Aug. 2016, pp. 785–794. <https://doi.org/10.1145/2939672.2939785>.
34. T. Kluyver et al., 'Jupyter Notebooks - a publishing format for reproducible computational workflows', in *International Conference on Electronic Publishing*, 2016. [Online]. Available: <https://api.semanticscholar.org/CorpusID:36928206>
35. J. VanderPlas, *Python data science handbook: essential tools for working with data*, First edition. Beijing Boston Farnham Sebastopol Tokyo: O'Reilly, 2016.
36. L. Rokach and O. Maimon, *Data Mining with Decision Trees: Theory and Applications*, 2nd ed., vol. 81. in Series in Machine Perception and Artificial Intelligence, vol. 81. WORLD SCIENTIFIC, 2014. <https://doi.org/10.1142/9097>.

37. L. Breiman, 'Random Forests', *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001. <https://doi.org/10.1023/A:1010933404324>.
38. J. H. Friedman, 'Greedy function approximation: A gradient boosting machine.', *Ann. Stat.*, vol. 29, no. 5, Oct. 2001. <https://doi.org/10.1214/aos/1013203451>.
39. D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*, 1st ed. in Wiley Series in Probability and Statistics. Wiley, 2013. <https://doi.org/10.1002/9781118548387>.
40. D. M. W. Powers, 'Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation', 2020. <https://doi.org/10.48550/ARXIV.2010.16061>.
41. M. T. Ribeiro, S. Singh, and C. Guestrin, '"Why Should I Trust You?": Explaining the Predictions of Any Classifier', 2016, *arXiv*. <https://doi.org/10.48550/ARXIV.1602.04938>.
42. A. Al-Kaisy, Z. Kirkemo, D. Veneziano, and C. Dorrington, 'Traffic Use of Rest Areas on Rural Highways: Recent Empirical Study', *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2255, no. 1, pp. 146–155, Jan. 2011. <https://doi.org/10.3141/2255-16>.
43. S. Tanaka, S. Ohno, and F. Nakamura, 'Analysis on drivers' parking lot choice behaviors in expressway rest area', *Transp. Res. Procedia*, vol. 25, pp. 1342–1351, 2017. <https://doi.org/10.1016/j.trpro.2017.05.158>.
44. Y. Zhang, Z. Pan, F. Zhu, C. Shi, and X. Yang, 'Quantitative Estimation and Analysis of Spatiotemporal Delay Effects in Expressway Traffic Accidents', *ISPRS Int. J. Geo-Inf.*, vol. 13, no. 11, p. 407, Nov. 2024. <https://doi.org/10.3390/ijgi13110407>.
45. W. Ji *et al.*, 'Spatial and temporal distribution of expressway and its relationships to land cover and population: A case study of Beijing, China', *Transp. Res. Part Transp. Environ.*, vol. 32, pp. 86–96, Oct. 2014. <https://doi.org/10.1016/j.trd.2014.07.010>.
46. S. Gutmann, C. Maget, M. Spangler, and K. Bogenberger, 'Truck Parking Occupancy Prediction: XGBoost-LSTM Model Fusion', *Front. Future Transp.*, vol. 2, p. 693708, Jul. 2021. <https://doi.org/10.3389/ffutr.2021.693708>.
47. S. S. Channamallu, S. Kermanshachi, J. M. Rosenberger, and A. Pamidimukkala, 'Parking occupancy prediction and analysis - a comprehensive study', *Transp. Res. Procedia*, vol. 73, pp. 297–304, 2023. <https://doi.org/10.1016/j.trpro.2023.11.921>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.