

A Random Forest Based Spatiotemporal Model for Parking Revenue Potential Classification to Support Adaptive Progressive Tariff Policies

Siti Rhofiah

Informatics Engineering Study Program, Telkom University, Purwokerto, Indonesia
E-mail: sitirhofiah@student.telkomuniversity.ac.id
ORCID iD: <https://orcid.org/0009-0007-7061-9369>

Sudianto Sudianto*

Informatics Engineering Study Program, Telkom University, Purwokerto, Indonesia
E-mail: sudianto@telkomuniversity.ac.id
ORCID iD: <https://orcid.org/0000-0001-8610-0587>

*Corresponding author

Aminatus Sa'adah

Informatics Engineering Study Program, Telkom University, Purwokerto, Indonesia
E-mail: aminatuss@telkomuniversity.ac.id
ORCID iD: <https://orcid.org/0000-0002-3469-7745>

Received: January 22, 2026; Revised: February 25, 2026; Accepted: March 24, 2026; Published: August 8, 2026

Abstract: The rapid growth of motor vehicles in urban areas has led to an increasing demand for parking facilities and requires tariff policies that are more adaptive to actual field conditions. This study aims to develop a parking revenue potential classification model to support the formulation of progressive parking tariff policies using a data-driven and spatial analysis approach. The dataset includes vehicle attributes, parking volume, and temporal parking usage patterns. Parking revenue potential is categorized into low, medium, and high classes using a quantile-based approach. Unlike most previous studies that focus on parking occupancy prediction, this study proposes a revenue-oriented spatiotemporal classification model integrating spatial coordinates and temporal parking patterns to support adaptive tariff policy formulation. The Random Forest algorithm is applied to classify parking revenue potential into low, medium, and high categories, achieving an accuracy of 90.24% for two-wheeled vehicles and 89.02% for four-wheeled vehicles. The classification results are integrated into an interactive Streamlit-based dashboard that visualizes the spatial distribution of parking revenue potential and enables simulations of progressive tariff adjustments based on spatial zones and temporal conditions. The proposed system functions as a decision support tool for parking management, aiming to improve operational efficiency and sustainably increase regional parking revenue.

Index Terms: Parking, Progressive Tariff, Random Forest, Classification, Spatial Analysis, Interactive Dashboard.

1. Introduction

Urban parking management has become a critical issue in many cities due to increasing vehicle ownership and limited parking space availability. Ineffective parking control may lead to illegal parking practices that disrupt traffic flow and reduce urban sustainability [1]. In the Indonesian context, parking management is closely related to local government revenue through parking retribution schemes, which are regulated by regional policies [2, 3].

Several studies have examined parking retribution and tariff policies as instruments to increase local revenue while maintaining public acceptance. Progressive on-street parking tariffs based on parking duration have been proposed as an effective approach to regulate parking demand in urban areas [4]. Policy evaluations also show that parking tariff increases must be supported by clear regulatory frameworks to ensure effectiveness and fairness [5].

Comprehensive parking management strategies emphasize not only pricing mechanisms but also demand management and system efficiency. Litman [6] highlighted that parking pricing plays a strategic role in influencing user

behavior and optimizing parking utilization. Progressive pricing strategies have further been discussed as a tool to manage parking demand and improve revenue performance [7, 8].

Recent advancements in data analytics and machine learning have enabled more accurate modeling of parking demand and user behavior. Machine learning approaches have been applied to predict parking demand and analyze the impact of parking fees on parking behavior [9]. Spatial and temporal characteristics of parking occupancy have also been explored using clustering and predictive modeling techniques [10, 11]. These developments indicate the potential of data-driven approaches to support more adaptive and efficient parking management policies. Although previous studies have investigated parking demand prediction and pricing mechanisms, limited research focuses on parking revenue potential classification integrating spatial attributes and machine learning techniques for policy-oriented tariff design. Therefore, this study proposes a Random Forest-based spatiotemporal classification model to support adaptive progressive parking tariff policies.

2. Related Work

Smart parking systems have been widely studied as part of intelligent transportation systems through the integration of Internet of Things (IoT), data analytics, and machine learning technologies. Channamallu et al. [12] conducted a comparative study of parking occupancy prediction models and demonstrated that machine learning-based approaches outperform traditional statistical methods. Similarly, Wang et al. [13] highlighted the importance of predictive analytics in optimizing revenue within smart parking ecosystems. However, these studies mainly emphasize parking occupancy prediction and operational efficiency, while limited attention has been given to revenue-oriented classification models that support tariff policy formulation.

Machine learning and deep learning techniques have been extensively applied to parking availability and demand prediction. Awan et al. [14] compared various machine and deep learning models and found that ensemble-based methods provide robust performance across different parking scenarios. Deep learning approaches have also been employed to classify traffic and parking patterns in smart city environments, particularly for handling complex spatiotemporal data [15]. Despite their effectiveness in prediction tasks, these approaches generally focus on parking availability forecasting rather than analyzing the potential revenue distribution of parking locations.

Among ensemble learning methods, Random Forest has gained significant attention due to its robustness against overfitting and its ability to handle heterogeneous features. Chen et al. [16] provided a comprehensive guide on the application of CART and Random Forest models in decision-making problems. Random Forest has also been successfully applied in applied classification tasks using structured datasets, further supporting its suitability for practical data-driven analysis [17]. Nevertheless, the application of Random Forest specifically for spatially informed parking revenue classification remains relatively underexplored.

Machine learning approaches have also been utilized in broader transportation studies to model temporal dynamics and mobility patterns. Li et al. [18] emphasized the role of temporal predictability in transportation systems, highlighting the importance of incorporating time-based features into analytical models.

Adaptive and dynamic pricing mechanisms have been explored as demand management tools in transportation systems. Valogianni et al. [19] investigated adaptive pricing strategies to regulate user behavior and system efficiency. In the context of smart parking, IoT-enabled architectures combined with advanced analytics have been proposed to support intelligent parking management [15, 20].

Despite the growing body of research on smart parking systems, most existing studies primarily focus on parking occupancy, availability, or demand prediction rather than revenue-oriented classification [12, 14, 15]. Furthermore, the explicit integration of spatial attributes such as parking location coordinates into machine learning-based parking revenue classification models remains limited [18, 21]. Therefore, this study addresses these gaps by proposing a Random Forest-based spatiotemporal classification framework to support adaptive and progressive parking tariff policy formulation. Recent studies increasingly explore the integration of machine learning and Internet of Things (IoT) technologies for data-driven parking management in smart city environments [22]. Advanced frameworks such as digital twin systems and deep learning models have also been proposed to improve parking demand prediction and urban mobility optimization [23, 24]. These developments highlight the growing role of data-driven analytics in supporting more adaptive and evidence-based parking management policies.

3. Materials and Methods

This research adopts a quantitative methodology integrating Data Mining and Spatial Analysis approaches. The overall workflow of this study is illustrated in Fig. 1, follows a systematic sequence of four main phases: (1) Data Collection, (2) Data Preprocessing, (3) Model Development and Evaluation, (4) Dynamic Progressive Pricing, and (5) Spatial Analysis Dashboard Streamlit Visualization.



Fig. 1. Research workflow of the adaptive parking tariff system.

3.1. Data Collection

The dataset used in this study consists of both secondary and primary data. The secondary data were obtained from the official report archives of the *Badan Perencanaan Pembangunan, Penelitian, dan Pengembangan Daerah* (BAPPEDALITBANG, the Regional Development Planning, Research, and Innovation Agency) of Banyumas Regency. The data, recorded in 2023, were collected as part of the regional parking retribution management and fiscal evaluation program. These data include information on parking revenue, vehicle volume, and operating time for motorcycles and cars across several urban parking locations.

In addition, primary data were collected by the researcher through direct field observation to obtain the spatial coordinates (latitude and longitude) of each parking point. The coordinates were initially recorded manually based on the identified parking locations. To minimize potential positional inaccuracies, the recorded coordinates were cross-checked and validated using multiple digital mapping platforms, including Google Maps and OpenStreetMap. This validation step helps ensure spatial consistency and improves the positional accuracy of the dataset for geospatial and machine learning analysis.

A total of 407 data records were compiled, each representing one parking location within the Banyumas Regency area. Although the dataset consists of 407 records, it represents official parking locations recorded by the local government. Therefore, each observation reflects an actual operational parking point, providing realistic insights into urban parking revenue distribution. The dataset was obtained from official records provided by the Regional Development Planning Agency (BAPPEDALITBANG), ensuring data reliability and authenticity.

To enhance model interpretability, the original attributes from the BAPPEDALITBANG report were refined and transformed into specific independent and dependent variables used in the analysis, as summarized in Table 1.

Table 1. Summary of Research Variables.

Type of Variable	Variable Name	Description
Independent Variables (Features)	Spatial Data	Latitude, Longitude, and Parking Point Name
	Quantitative Data	Vehicle count (motorcycles and cars) for weekday and weekend.
	Temporal Data	Peak, moderate, and off-peak hours for weekdays and weekends (decimal format).
	Revenue Data	Annual parking revenue per vehicle type.
Dependent Variable (Target)	Tariff Potential Class	Categorical classification (Low, Medium, High) based on annual revenue quantiles.

By combining spatial and fiscal data, the resulting dataset supports both machine learning-based prediction and spatial analysis of parking tariff potential. The refined structure ensures that each observation contains synchronized spatial, temporal, and financial information necessary for adaptive policy modeling.

3.2. Data Preprocessing and Feature Engineering

This stage aims to transform the raw dataset into a clean and structured format suitable for processing by the Machine Learning model. Several preprocessing techniques were applied to improve data consistency, accuracy, and readiness for model training.

3.2.1. Data Cleaning

1. Data Encoding: Revenue values originally stored in text format (e.g., “IDR 150,000”) were converted into numerical decimal form (e.g., 150000.0) to ensure computational compatibility within the model.
2. Time Data Conversion: Time-range attributes (e.g., “15.00–17.00”) were transformed into single average decimal hour values (e.g., 16.0) using a custom `converttimerange()` function. The resulting values were utilized as quantitative features in the model.

3. **Missing Value Imputation:** Missing values (NaN) detected in numerical attributes such as vehicle quantity, revenue, and time (decimal hours) were imputed using median or mean substitution, depending on the variable's distribution. This step ensured that no critical records were removed during model training, thereby maintaining the dataset's statistical integrity and sample completeness.

After performing the data cleaning and imputation process, a total of 405 valid data records remained from the initial 407. Two records were removed due to incomplete or inconsistent attribute values that could not be corrected during preprocessing.

3.2.2. Data Splitting

Following data cleaning, the dataset was divided into training and testing subsets using an 80:20 ratio to enable unbiased model evaluation. The `train_test_split()` function from the scikit-learn library was utilized to ensure random yet reproducible sampling. Consequently, 324 records were allocated for model training and 81 records for testing. This approach maintains data representativeness while preventing overfitting during the training phase. Although k-fold cross-validation is commonly used to obtain more robust model evaluation, this study adopted a train-test split strategy due to the moderate dataset size and the need to preserve sufficient samples for model training. This approach ensures an independent testing subset while maintaining data representativeness. Future studies may explore k-fold cross-validation to further improve model robustness and reduce potential overfitting risks.

3.2.3. Feature Engineering

1. **Total Revenue Aggregation:** New columns, `TotalRevenueMotorcycle` and `TotalRevenueCar`, were generated by aggregating the annual revenue data from respective vehicle categories.
2. **Formation of Potential Class (Target Variable):** The total annual revenue was transformed into categorical labels using the Quantile Binning technique (`pd.qcut`). The resulting classes represent the parking point's revenue potential:
 - a) **Low Class:** First quantile (lowest 33% of revenue);
 - b) **Medium Class:** Between the first and third quantiles (middle 34%);
 - c) **High Class:** Third quantile and above (highest 33% of revenue).

Although preprocessing techniques such as mean/median imputation and quantile-based binning are commonly used in machine learning pipelines, their use in this study aims to improve data reliability and enhance model interpretability. Median imputation was selected because it is more robust to extreme values than mean substitution, particularly for revenue-related attributes.

Meanwhile, quantile-based binning transforms continuous revenue values into categorical classes (Low, Medium, High), allowing the Random Forest classifier to capture relative differences in revenue potential among parking locations. This transformation also helps reduce noise in the dataset and improves the model's ability to identify meaningful relationships between spatial, temporal, and vehicle-volume features.

Furthermore, the quantile-based labeling strategy was adopted to mitigate potential class imbalance during model training. Rather than representing absolute economic thresholds, the resulting classes reflect the relative revenue performance of parking locations within the study area. This relative classification enables policymakers to identify low-, medium-, and high-potential parking zones to support adaptive parking tariff policies while maintaining stable model performance.

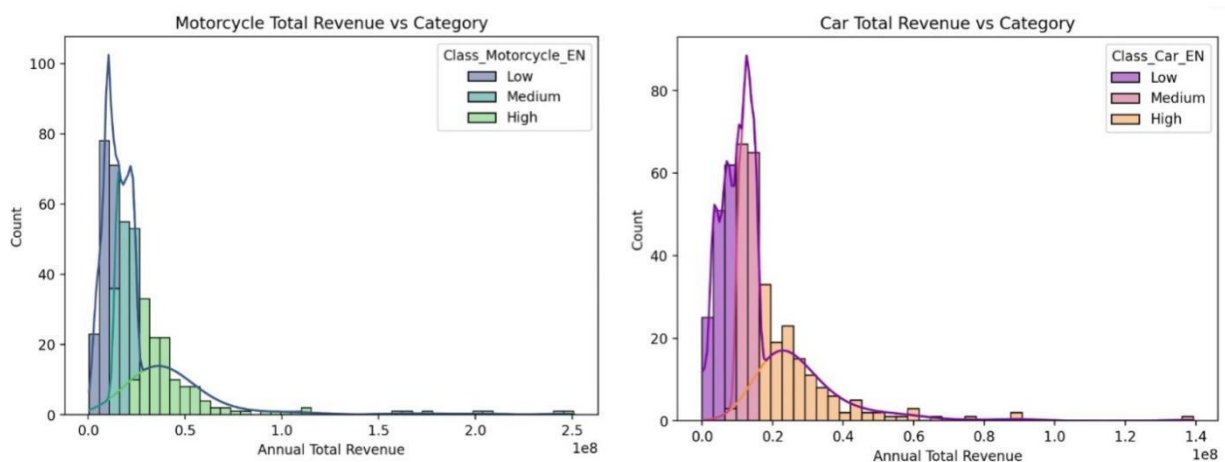


Fig. 2. Distribution of Annual Total Revenue by Potential Class for Motorcycles and Cars.

The resulting distribution of total annual parking revenue for motorcycles and cars after quantile-based categorization is illustrated in Fig. 2. The histogram and KDE (Kernel Density Estimation) plots show the skewness and separability of revenue across the three classes, confirming that quantile binning effectively partitions the data into balanced revenue-based segments for model training.

3.3. Random Forest Classification Model

3.3.1. Model Structure and Working Principle

The Random Forest Classifier (RFC) constructs multiple decision trees and aggregates their outputs through a majority-voting mechanism to enhance predictive accuracy and minimize overfitting. Each decision tree is trained on a random subset of features and data samples, thereby promoting model diversity. This ensemble approach improves generalization by reducing variance across individual trees.

The majority-voting principle is mathematically expressed in Equation (1), where the final class prediction $\hat{y}$ is determined by the most frequent class output among T individual decision trees:

$$\hat{y} = \text{mode}\{h_1(x), h_2(x), \dots, h_t(x)\} \quad (1)$$

where $\hat{y}$ represents the final prediction result, mode refers to the prediction value that appears most frequently, and $h_t(x)$ is the prediction produced by the t -th decision tree. Each tree generates its prediction independently, and the final classification result is determined through a majority voting mechanism, which selects the class that appears most frequently among all decision trees.

3.3.2. Gini Impurity and Node Splitting Criterion

The optimal split at each node is determined using the *Gini Impurity*, which measures the degree of class heterogeneity within a node. The Gini impurity for a node S_i and its corresponding split are calculated as shown in Equation (2) and Equation (3).

$$\text{Gini}(S_i) = 1 - \sum_{j=1}^c p_j^2 \quad (2)$$

$$\text{Gini}_{\text{split}} = \sum_{i=1}^k \frac{n_i}{n} \times \text{Gini}(S_i) \quad (3)$$

where p_j denotes the probability of class j in node S_i , n_i is the number of samples in subset S_i , and n represents the total number of samples in the parent node. A lower Gini value indicates a purer node and therefore a better feature split.

3.3.3. Model Training and Evaluation

The Random Forest model was tuned using several hyperparameters summarized in Table 2. The modeling process was conducted systematically to ensure robustness and reliability in the classification outcomes. The dataset was divided into training and testing subsets using an 80:20 ratio, where the training data were used to construct the model and the testing data to evaluate generalization performance on unseen samples. The hyperparameter configuration was selected based on commonly adopted settings in ensemble learning literature and preliminary experimentation on the dataset. The number of trees ($n_{\text{estimators}} = 150$) was chosen to ensure stable ensemble voting while maintaining computational efficiency. The maximum tree depth was limited to 15 to capture nonlinear relationships while preventing excessive model complexity that could lead to overfitting. These parameter choices provide a balance between predictive performance and model generalization for the dataset size used in this study. Although k-fold cross-validation is commonly used for model validation, a train-test split strategy was adopted in this study to preserve a sufficient number of samples for model training while maintaining an independent dataset for evaluation. Considering the moderate dataset size, this approach provides a practical compromise between training stability and unbiased performance assessment.

Subsequently, the Random Forest Classifier was trained separately for each vehicle category motorcycles and cars, using predictive features such as vehicle quantity and average parking duration. These features were employed to predict the corresponding target variables, *Class_Motorcycle* and *Class_Car*, which represent the parking tariff potential categories: *Low*, *Medium*, and *High*.

3.3.4. Performance Metrics

After model training, performance was evaluated using standard classification metrics derived from the confusion matrix: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). The evaluation metrics are defined as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (7)$$

As shown in Equation (4), Accuracy measures the overall proportion of correctly classified instances, while Precision and Recall, defined in Equations (5) and (6), evaluate the model's exactness and completeness in identifying the correct class. The F1-Score, expressed in Equation (7), represents the harmonic mean of Precision and Recall, providing a balanced performance indicator that is particularly useful in datasets with class imbalance.

All modeling and evaluation processes were implemented using the scikit-learn Python library, ensuring reproducibility and methodological consistency.

Table 2. Random Forest Hyperparameters.

Hyperparameter	Value	Description
n_estimators	150	Number of trees for stable voting and generalization
max_depth	15	Maximum tree depth to capture complex relationships
min_samples_split	2	Minimum samples required to split an internal node
min_samples_leaf	3	Minimum samples required at a leaf node
bootstrap	TRUE	Enables bootstrap sampling to improve robustness
random_state	42	Ensures reproducibility
criterion	Gini	Gini impurity as split quality function

3.4. Adaptive and Dynamic Progressive Tariff Scheme

The predicted parking potential class is used as input for an adaptive pricing mechanism, where each class is assigned an adaptive base tariff and a dynamic surcharge applied during peak hours. This strategy aims to regulate parking demand while simultaneously maximizing parking revenue. The classification result (potentialclass) is directly used as an input for the adaptive tariff function, making the system both dynamic and progressive.

3.4.1. Adaptive Base Tariff Allocation

The adaptive base tariff scheme for each vehicle type and potential class is presented in Table 3. High-potential parking locations are assigned higher base tariffs to reflect their greater demand intensity, whereas low-potential locations implement lower tariffs to maintain accessibility and ensure an equitable distribution of parking demand. This adaptive structure allows tariff differentiation based on real spatial revenue potential, forming a data-driven foundation for equitable and efficient pricing policy.

3.4.2. Dynamic Progressive Surcharge

To further regulate parking behavior during high-demand periods, a dynamic progressive surcharge is applied during peak hours, as summarized in Table 4. High-potential areas receive the highest surcharge rates during peak periods, while low-potential zones are exempted from additional charges. This time-dependent mechanism provides flexibility for demand management, discouraging prolonged vehicle occupancy in high-demand locations.

Table 3. Adaptive Base Tariff Scheme.

Vehicle Type	Potential Class	Adaptive Base Tariff (IDR)
Motorcycle	High	3,000
Motorcycle	Medium	2,000
Motorcycle	Low	1,000
Car	High	5,000
Car	Medium	4,000
Car	Low	3,000

Table 4. Dynamic Progressive Surcharge System.

Potential Class	Time Condition (Hour > 9.0)	Surcharge (IDR)
High	Yes	1,000
Medium	Yes	500
Low	Yes	0

3.4.3. Final Tariff Formulation

The overall adaptive tariff formulation is expressed in Equation (8):

$$\text{Final Tariff} = \text{Adaptive Base Tariff} + \text{Progressive Surcharge} \quad (8)$$

In this formulation, the *Adaptive Base Tariff* corresponds to the tariff value assigned based on the predicted revenue potential class (e.g., High: IDR 3,000; Medium: IDR 2,000; Low: IDR 1,000), while the *Progressive Surcharge* represents an additional fee imposed during predefined peak hours (e.g., after 09:00 AM). The surcharge value also adapts according to the location's potential class, ensuring a more aggressive price increase for high-demand areas. This adaptive and dynamic tariff framework provides a data-driven foundation for balancing accessibility, regulating vehicle flow, and supporting fiscal optimization in urban parking management.

3.5. Spatial Visualization and Dashboard

To support tariff simulation and temporal demand analysis, parking activity is categorized into three time-based demand levels, namely Off-Peak, Moderate, and Peak hours. The classification rules for each time category are defined based on daily parking activity patterns and are summarized in Table 5. These time categories are used as inputs for the dynamic pricing simulation module in the dashboard.

Table 5. Time Category Rules for Simulation.

Category	Time Range	Decimal Hour	Demand Characteristics
Off-Peak	00:00–06:00, 22:00–24:00	hour ≤ 6 or hour ≥ 22	Low demand, minimal vehicles, night rest period
Peak	09:00–19:00	9 < hour ≤ 19	High demand, heavy vehicle volume, maximum activity period
Moderate	06:00–09:00, 19:00–22:00	6 < hour ≤ 9 or 19 < hour < 22	Transition demand, medium vehicle volume, morning and evening peak hours

4. Results and Discussions

4.1. Classification of Parking Revenue Potential Based on Quantile Thresholds

The target variable, namely the *Parking Revenue Potential Class* (Low, Medium, High), was established using a quantile-based approach applied to the annual parking revenue distribution. This method ensures that each class represents an equal proportion of the empirical data, thereby reflecting realistic revenue stratification across parking locations. Table 6 presents the resulting revenue thresholds and the corresponding suggested base tariffs for motorcycles and cars.

For motorcycles, the low potential class corresponds to locations generating less than IDR 13,344,000 annually, while high potential locations exceed IDR 25,075,008. In contrast, car parking locations exhibit lower revenue thresholds due to lower parking frequency, with high potential locations exceeding IDR 16,268,160 annually.

These thresholds provide a quantitative foundation for adaptive tariff zoning, ensuring that base tariffs are aligned with actual parking demand and revenue generation capacity at each location.

Table 6. Classification of Parking Revenue Potential Based on Quantile Thresholds.

Vehicle Type	Potential Class	Actual Annual Revenue Threshold (IDR)	Suggested Initial Base Price (IDR)
Motorcycle	Low	Revenue <13,344,000	1,000
	Medium	13,344,000 – 25,075,008	2,000
	High	Revenue >25,075,008	3,000
Car	Low	Revenue <9,587,040	3,000
	Medium	9,587,040 – 16,268,160	4,000
	High	Revenue >16,268,160	5,000

Note: Suggested base prices are indicative values for adaptive tariff policy.

4.2. Random Forest Model Performance

The Random Forest classifier demonstrates strong predictive performance for both vehicle categories. The classification accuracy achieved is 90.24% for motorcycles and 89.02% for cars, indicating reliable generalization on unseen test data. Although the obtained accuracy values appear relatively high, this performance can be explained by the strong relationship between vehicle volume, temporal parking patterns, and revenue generation. These variables provide highly informative signals for distinguishing revenue potential levels across parking locations. Therefore, the model is able to learn clear patterns in the dataset without requiring highly complex feature transformations.

Table 7 summarizes the detailed classification metrics for each class. For motorcycle parking, the model achieves an F1-score of 0.96 for the High potential class, indicating excellent discrimination of high-revenue locations. Similarly, the Medium and Low classes show balanced precision and recall values, confirming the robustness of the model across all revenue segments.

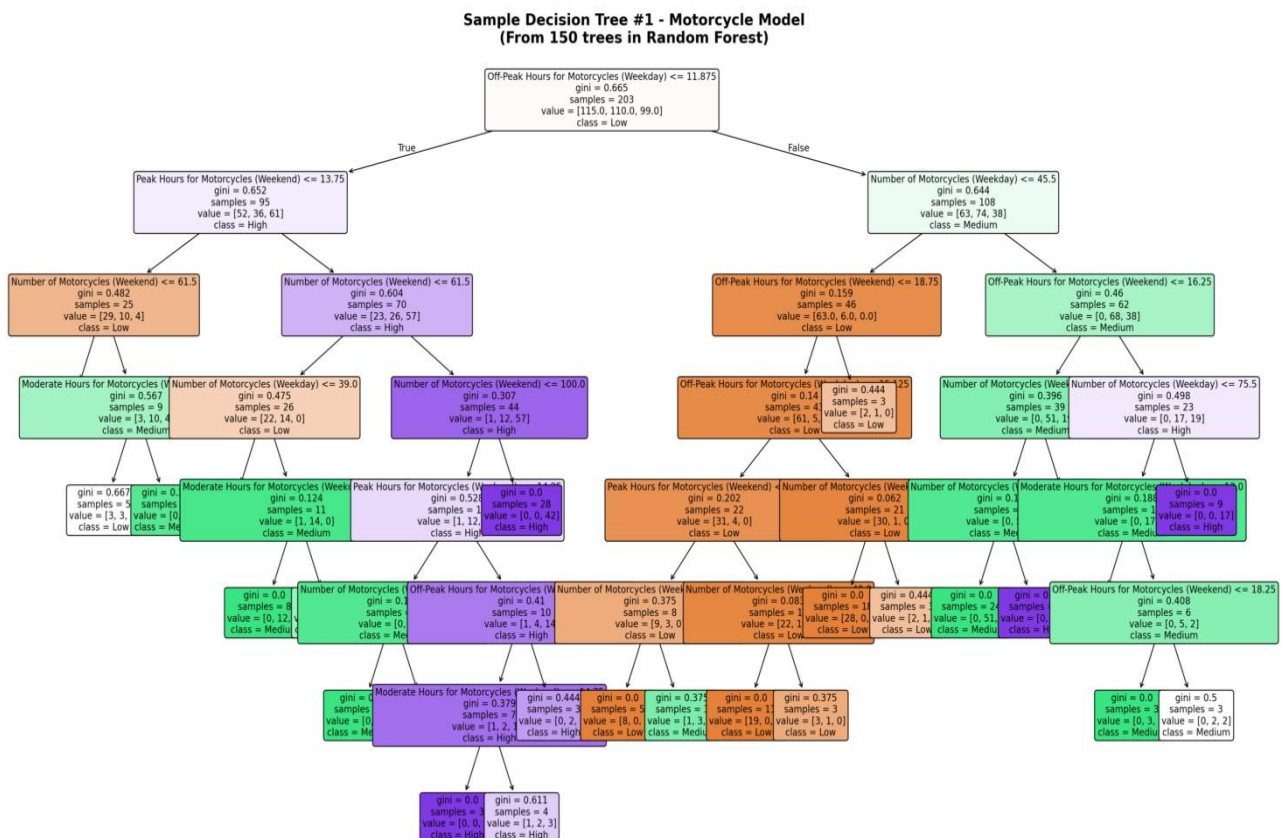
For car parking, the High potential class also exhibits strong predictive reliability with an F1-score above 0.91. The slightly lower performance for the Medium class reflects the overlapping revenue characteristics among moderate-demand locations, which is common in real-world parking behavior patterns.

The learning curve shown in Fig. 5 indicates that model accuracy stabilizes after approximately 40 trees, confirming that the selected hyperparameter configuration achieves a good balance between bias and variance. Meanwhile, the decision tree visualizations in Figures 3 and 4 illustrate how the Random Forest model captures complex nonlinear decision boundaries based on traffic volume and temporal demand features.

4.2.1. Feature Importance Analysis

To enhance model interpretability, feature importance analysis was conducted to evaluate the contribution of each predictor to the classification process. Fig. 6 presents the five most influential features for cars and motorcycles, respectively. Overall, vehicle volume is the most decisive factor for both vehicle types, while temporal variables refine classification. This analysis measured the relative contribution of each variable to the model's classification. Fig. 6 shows the five most influential features for cars and motorcycles, respectively.

- Cars: Number of Cars (Weekday) = 0.4887, Number of Cars (Weekend) = 0.3544, with temporal variables (OffPeak, Moderate Hours) around 0.02–0.03.
- Motorcycles: Number of Motorcycles (Weekday) = 0.4931, Number of Motorcycles (Weekend) = 0.3422, temporal factors (Off-Peak, Peak Hours) around 0.026–0.033.



These findings confirm that parking volume is the primary determinant of revenue potential, while temporal patterns act as complementary factors that improve classification precision. Although the dominance of vehicle volume may appear intuitive, the model provides quantitative evidence confirming its influence on parking revenue potential. This empirical validation is important because it supports transportation demand management theories that link parking demand intensity with pricing and revenue generation. Furthermore, the analysis demonstrates how temporal variables complement vehicle volume in refining the classification boundaries.

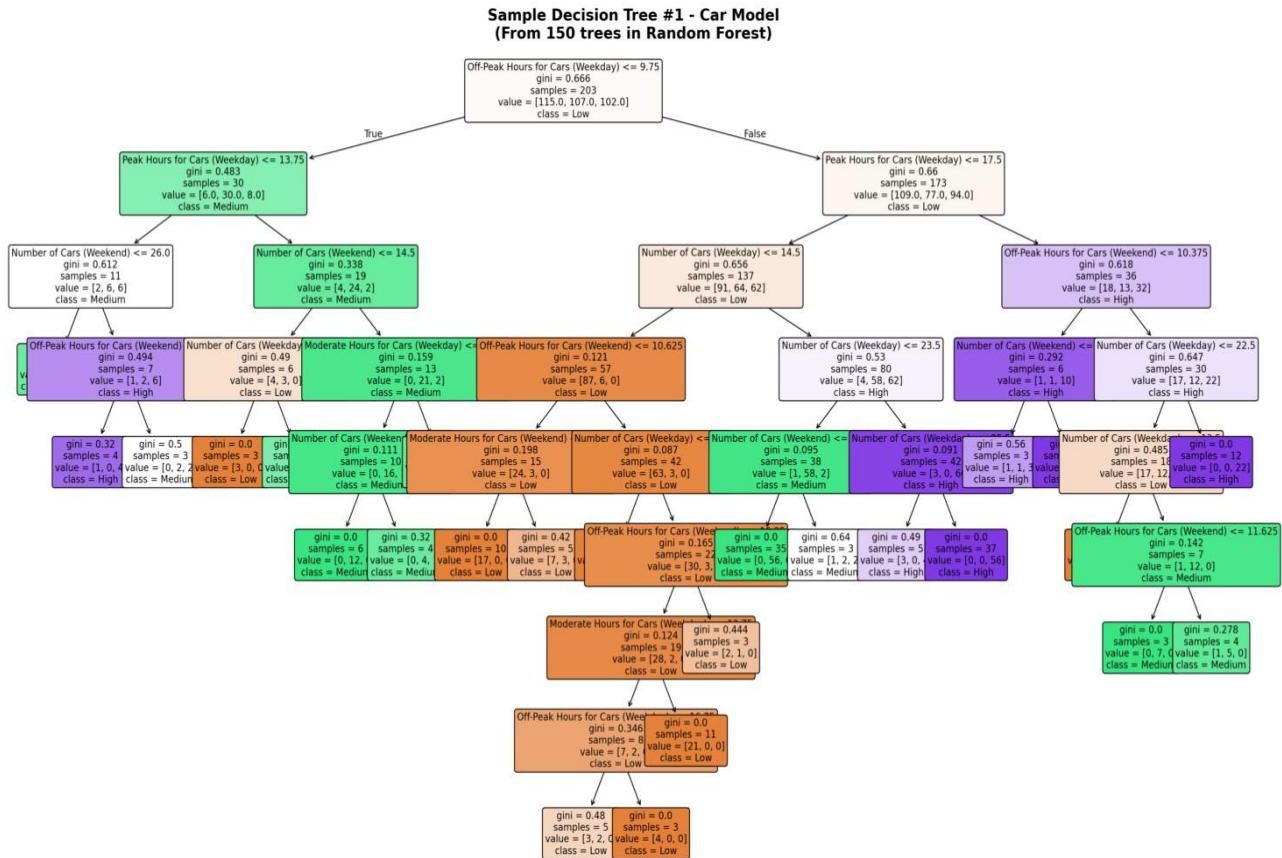


Fig. 4. Decision Tree for Car Classification.

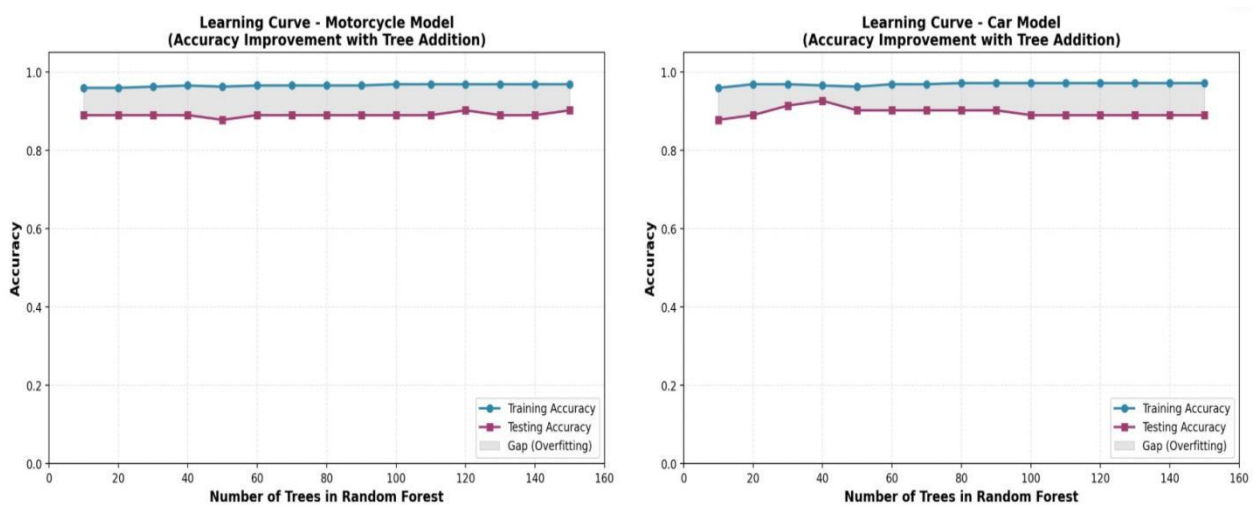


Fig. 5. Learning Curve of Random Forest Model.

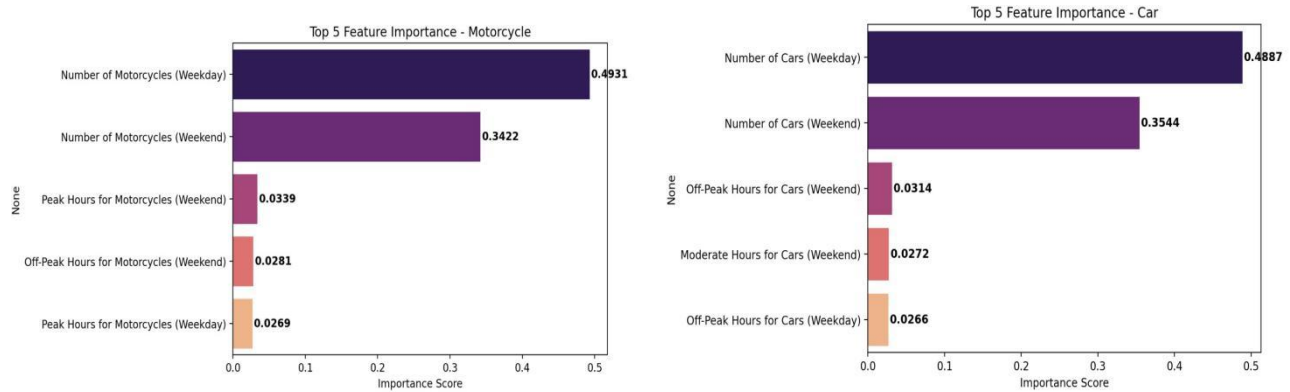


Fig. 6. Feature importance for motorcycles and cars.

4.2.2. Evaluation Confusion Matrix

The Confusion Matrix Visualization Fig. 7 for motorcycles and cars confirm that most predictions align with actual labels, especially for the High and Medium classes, indicating that the Random Forest model effectively captures the variance of spatial and temporal factors influencing parking revenue potential.

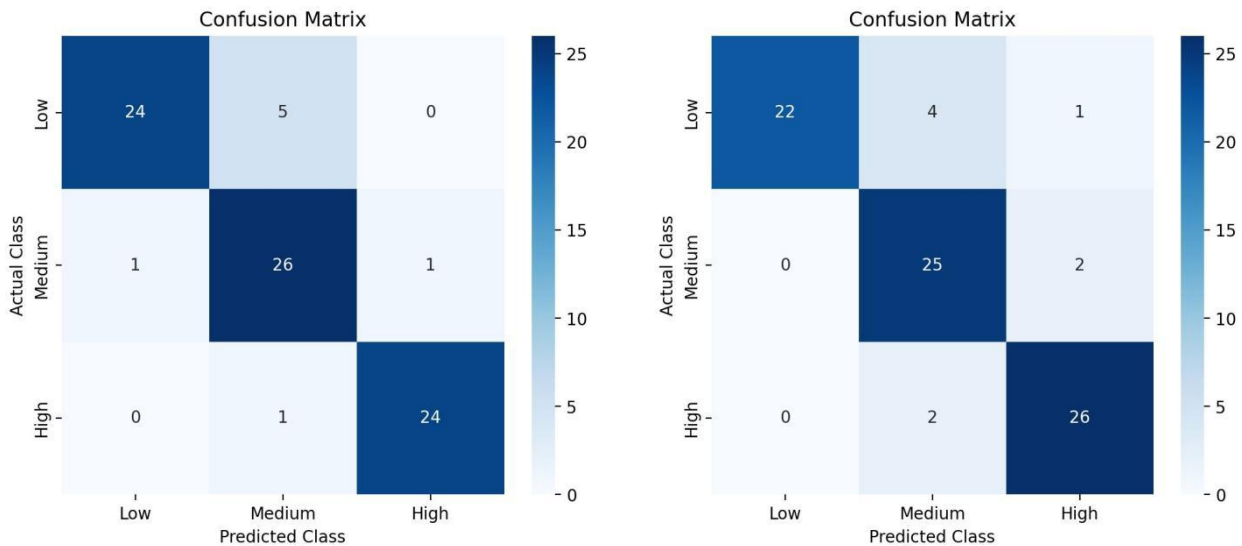


Fig. 7. Confusion Matrices for Motorcycle and Car Classification.

Table 7. Random Forest Model Performance.

Vehicle Type	Accuracy	Class	Precision (Weighted)	Recall (Weighted)	F1-Score (Weighted)
Motorcycle	0.90	High	0.96	0.96	0.96
		Medium	0.8125	0.9286	0.8667
		Low	0.96	0.8276	0.8889
Car	0.89	High	0.8966	0.9286	0.9123
		Medium	0.8065	0.9259	0.8621
		Low	1	0.8148	0.898

4.3. Visualization and System Implementation (Streamlit Dashboard)

The research outcomes were integrated into an interactive web-based dashboard developed using Streamlit and publicly accessible at: <https://parkingtariffprogressivebanyumas.streamlit.app/>. The dashboard serves as a proof-of-concept decision support system that combines machine learning classification, spatial visualization, and dynamic tariff simulation.

A key feature of this system is its interactive spatial visualization. Each parking location is georeferenced with latitude and longitude coordinates, allowing users to explore the spatial distribution of parking revenue potential across the urban area. Users can click on individual parking points to view a popup containing comprehensive information for that location, including:

- Predicted revenue potential class for both motorcycles and cars (Low, Medium, High).
- Base tariff assigned according to the classification and any adaptive progressive surcharge applied during peak hours.
- Location name for easy identification.
- Exact coordinates (latitude and longitude) for spatial reference.

Fig. 8 illustrates the main dashboard interface, which consists of four integrated modules:

- Data Table Module:** Displays raw and preprocessed datasets along with the assigned revenue potential classes.
- Visualization Module:** Provides descriptive analytics, including revenue histograms, quantile thresholds, and 24-hour parking demand trends.
- Modeling Module:** Presents evaluation metrics such as accuracy, confusion matrices, and feature importance scores.
- Map & Simulation Module:** Visualizes parking locations spatially on an interactive map. Each point is color-coded based on the predicted potential class and enables real-time tariff simulation based on temporal conditions. Fig. 9 and Fig. 10 present the spatial visualization and tariff simulation interfaces.

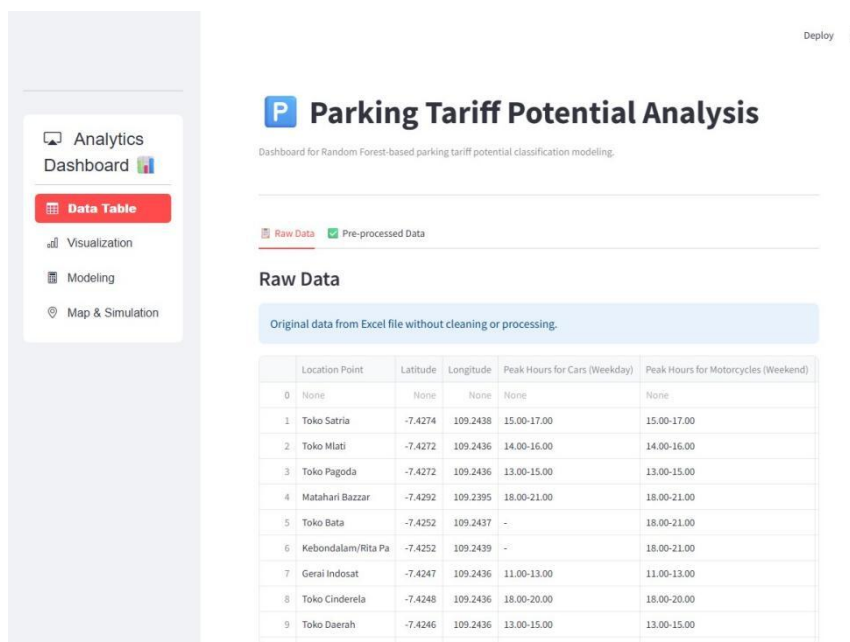


Fig. 8. Main dashboard interface.

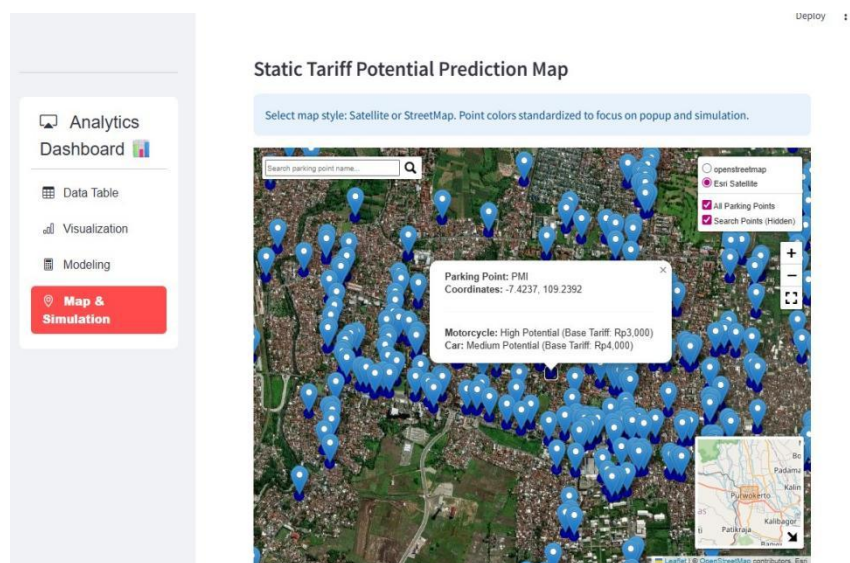


Fig. 9. Spatial visualization module.

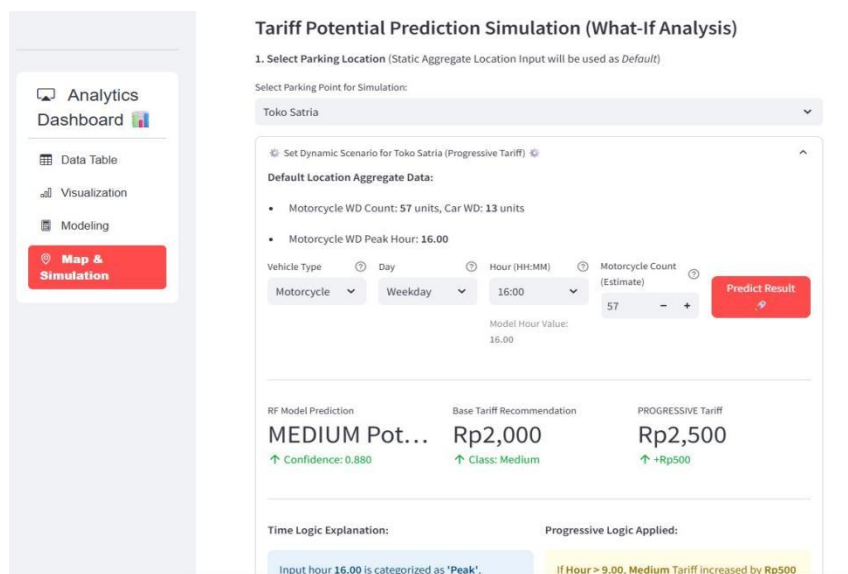


Fig. 10. Tariff simulation module.

This interactive spatial mapping functionality demonstrates how the proposed spatiotemporal classification model can be operationalized into a transparent, geographically adaptive, and evidence-based parking pricing mechanism, supporting both operational management and informed policy decisions.

4.4. Discussion

The analysis reveals a significant disparity between motorcycle and car revenue thresholds (Table 6). Despite lower per-unit tariffs, the higher frequency and volume of motorcycle parking result in substantially greater total revenues. This finding highlights the dominance of motorcycles in urban mobility patterns across Banyumas Regency and underscores the necessity for differentiated pricing policies based on vehicle type. The implication is that motorcycles, while individually low-cost, collectively contribute a major share of total parking income, reinforcing the rationale for adaptive and proportional tariff structures.

Model validation results further confirm the robustness of the Random Forest classifier, achieving accuracy levels between 0.89 and 0.90 for both vehicle categories. These high performance metrics demonstrate the reliability of the model in classifying parking tariff potential across spatial zones. The results also indicate that vehicle volume is the most influential predictor, aligning with the principles of Transportation Demand Management (TDM), which emphasize parking density and temporal demand as key determinants of turnover rate and pricing elasticity. Such consistency between model outcomes and established theoretical frameworks validates the suitability of the Random Forest approach for fiscal and spatial policy modeling in urban contexts.

The implementation of an adaptive and dynamic progressive pricing system further strengthens this analysis. The proposed framework integrates spatial adaptivity and temporal progressivity within a single pricing logic. Spatial adaptivity ensures that each parking location is assigned a base tariff according to its predicted potential class (Low, Medium, High), forming an Adaptive Spatial Tariff Map (Fig. 9). Meanwhile, temporal progressivity increases tariffs progressively during peak hours (e.g., after 09:00 AM) in proportion to the location's potential class. For instance, in a simulation at Toko Satria, under conditions of "Motorcycle–Weekday–16:00–100 vehicles," the model predicted a Medium Potential class with 0.98 confidence, resulting in a base tariff of IDR 2,000, a progressive surcharge of IDR 500, and a final tariff of IDR 2,500. This two-dimensional pricing mechanism enables real-time demand management by encouraging turnover in high-demand areas while maintaining accessibility and equity in lower-demand zones.

From a spatial policy perspective, the study provides a quantitative and integrated framework for local governments to design evidence-based parking tariff systems. By mapping the classification results (High, Medium, Low Potential) defined by the actual quantile thresholds (Table 6), policymakers can precisely delineate tariff zones aligned with fiscal potential and demand intensity. This zoning approach promotes a smart and sustainable fiscal management strategy, grounded in empirical revenue distribution patterns and supported by adaptive algorithms. The resulting spatial classification map functions as an analytical foundation for defining location-based tariffs, aligning with TDM objectives to optimize both parking utilization and municipal revenue in a balanced, data-driven manner.

5. Conclusion and Recommendations

Based on the analysis and discussions presented in this study, it can be concluded that the research objectives have been successfully achieved. This research developed a Random Forest-based parking revenue potential classification model integrated with spatial analysis, serving as the foundation for formulating data-driven progressive parking tariff policies. The analysis effectively categorized parking locations based on vehicle type, usage volume, and temporal activity patterns. The application of quantile binning enabled the classification of parking points into three revenue potential classes: Low, Medium, and High, with a relatively balanced distribution. Spatial analysis further revealed that high-potential parking areas are concentrated in zones of high economic activity and urban mobility, while low-potential areas are predominantly located in residential zones. Temporally, the highest parking demand occurs between 09:00 and 17:00, indicating the necessity for time-based tariff adjustments.

The Random Forest model demonstrated strong predictive capability, achieving 90.24% accuracy for motorcycles and 89.02% for cars using 150 decision trees with the Gini Index criterion. Weighted precision, recall, and F1-scores remained consistently high across all classes, reflecting good generalization performance and minimal overfitting. Feature importance analysis highlighted that vehicle volume, particularly weekday and weekend activity, is the most influential factor in determining parking revenue potential, while temporal variables act as complementary predictors that refine classification outcomes. Furthermore, the classification results were integrated into an interactive spatial dashboard that allows for the geographic visualization of parking revenue potential. Through its dynamic pricing simulation module, the system can adapt parking tariffs according to both spatial potential class and temporal demand, such as implementing higher tariffs during peak hours for high-potential areas. This approach enables transparent, equitable, and data-driven tariff zoning to support effective and adaptive parking management policies.

For future development, several directions are recommended. The inclusion of additional spatial and contextual variables, such as traffic density, economic activity level, weather patterns, and seasonal variations, may enhance model robustness and policy responsiveness. Field validation of parking occupancy data is also essential to ensure the model's predictions align with real-world operational conditions. The integration of real-time IoT-based data such as automated parking sensors, e-ticketing systems, and mobile parking applications would further enable dynamic and continuous tariff adjustment based on actual demand patterns. Additionally, incorporating the developed dashboard into local government digital parking management systems would facilitate transparent and accountable implementation of adaptive progressive tariffs. Finally, future studies are encouraged to analyze user behavior and responses to dynamic pricing models through surveys or simulation-based behavioral analysis to ensure that pricing policies are both economically efficient and socially acceptable.

All the Declarations and Statements

Author Contributions Statement

Siti Rhofiah – Conceptualization of the research, methodology development, data collection and curation, data analysis, Random Forest model development, software and Streamlit dashboard implementation, and writing the original draft of the manuscript.

Sudianto Sudianto – Conceptualization, supervision, methodology validation, and writing (review and editing of the manuscript).

Aminatus Sa'adah – Formal analysis, provided guidance on the analytical model and research methodology, and manuscript review.

All authors have read and agreed to the published version of the manuscript.

Conflict of Interest Statement

The authors declare no conflict of interest.

Funding Declaration

This research received no external funding.

Data Availability Statement

The datasets generated and analyzed during the current study are not publicly available due to government data policy but are available from the corresponding author on reasonable request.

Ethical Declarations

This study does not involve human participants or animals.

Acknowledgments

The authors would like to express their sincere gratitude to the Regional Development Planning, Research, and Innovation Agency (BAPPEDALITBANG) of Banyumas Regency for providing access to the data used in this research.

Declaration of Generative AI in Scholarly Writing

The authors declare that generative AI tools were used solely to improve the readability and language quality of the manuscript. All scientific content, analysis, and interpretation were generated by the authors.

Abbreviations

The following abbreviations are used in this manuscript:

AI - Artificial Intelligence

ML - Machine Learning

RFC - Random Forest Classifier

IoT - Internet of Things

BAPPEDALITBANG - Badan Perencanaan Pembangunan, Penelitian, dan Pengembangan Daerah

Appendix A/B/C..., with appendix title

None.

References

- [1] E. Erfina, L. Lukman, H. Hardianti, and B. B. Budai, "Illegal Parking Control Strategy to Improve Sustainable Management of Urban Areas," *Journal of Governance and Public Policy*, vol. 12, no. 1, pp. 112–124, Feb. 2025.
- [2] L. P. Tribuani, B. A. Saebani, and A. Saptaji, "Tinjauan Siyasaah Maliyah terhadap Pelaksanaan Retribusi Parkir dan Kontribusinya Kepada Pendapatan Asli Daerah Berdasarkan Peraturan Daerah Nomor 04 Tahun 2021 Tentang Retribusi Parkir di Kabupaten Purwakarta," *R2J*, vol. 7, no. 4, 2025.
- [3] BAPPEDALITBANG Kabupaten Banyumas dan PT. Mitra Nusa Konsulindo, *Laporan Akhir Kajian Analisis Potensi Parkir Kabupaten Banyumas*, Banyumas, Indonesia, 2023.
- [4] R. B. H. , T. D. K. Irawan Rizki Dwi, "Studi Tarif Parkir On Street Progresif Kawasan Perkotaan Menengah Berdasarkan Durasi Parkir," vol. 12, pp. 118–127, Aug. 2023.
- [5] S. P. Asmara, "Analisis Kebijakan Peraturan Daerah Tentang Kenaikan Tarif Parkir Di Dki Jakarta," *PUBLIKA : Jurnal Ilmu Administrasi Publik*, vol. 6, no. 2, pp. 246–256, Oct. 2020.
- [6] T. A. Litman, "www.vtpi.org Info@vtpi.org 250-508-5150 Parking Management Comprehensive Implementation Guide Parking Management: Comprehensive Implementation Guide," 2007. [Online]. Available: www.vtpi.org
- [7] D. A. Ormelas, P. Y. Park, and M. J. Roorda, "Managing Parking with Progressive Pricing Mehdi Nourinejad," 2022. [Online]. Available: <https://ssrn.com/abstract=4007922>
- [8] S. H. Bayih and S. L. Tilahun, "Dynamic vehicle parking pricing. A review," 2024, *Wroclaw University of Science and Technology*.
- [9] A. Bokadia, M. A. Ahmed, R. Das, and S. Mallik, "Impact of Parking Fee on On-Street Night Parking Demand Prediction in Delhi using Machine Learning Approaches," *Transp. Res. Rec.*, vol. 2679, no. 7, pp. 859–876, Jul. 2025.
- [10] M. Pönkänen, H. Tenkanen, and M. Mladenović, "Spatial accessibility and transport inequity in Finland: Open source models and perspectives from planning practice," *Comput. Environ. Urban Syst.*, vol. 116, Mar. 2025.
- [11] H. Candra and A. Andrianingsih, "Spatial Analysis of Random Forest Classification Model for Availability Mapping of Sports Facilities in Jakarta," *Journal of Artificial Intelligence and Software Engineering (J-AISE)*, vol. 5, no. 1, p. 228, Mar. 2025.
- [12] S. S. Channamallu, V. K. Padavala, S. Kermanshachi, J. M. Rosenberger, and A. Pamidimukkala, "Examining parking occupancy prediction models: a comparative analysis," in *Transportation Research Procedia*, Elsevier B.V., 2023, pp. 281–288.
- [13] T. Wang, S. Li, W. Li, Q. Yuan, J. Chen, and X. Tang, "A Short-Term Parking Demand Prediction Framework Integrating Overall and Internal Information," *Sustainability (Switzerland)*, vol. 15, no. 9, May 2023.
- [14] F. M. Awan, Y. Saleem, R. Minerva, and N. Crespi, "A comparative analysis of machine/deep learning models for parking space availability prediction," *Sensors (Switzerland)*, vol. 20, no. 1, Jan. 2020.
- [15] A. Dahiya et al., "Machine Learning-Based Prediction of Parking Space Availability in IoT-Enabled Smart Parking Management Systems," *J. Adv. Transp.*, vol. 2024, 2024.
- [16] Y. Chen, K. C. Cheung, R. Z. Sun, and S. C. P. Yam, "A user guide of CART and random forests with applications in FinTech and InsurTech," *Japanese Journal of Statistics and Data Science*, vol. 7, no. 2, pp. 999–1038, Nov. 2024.
- [17] A. G. Ismaeel et al., "Traffic Pattern Classification in Smart Cities Using Deep Recurrent Neural Network," *Sustainability (Switzerland)*, vol. 15, no. 19, Oct. 2023.
- [18] H. Li, F. He, X. Lin, Y. Wang, and M. Li, "Travel time reliability measure based on predictability using the Lempel–Ziv algorithm," *Transp. Res. Part C Emerg. Technol.*, vol. 101, pp. 161–180, Apr. 2020.
- [19] K. Valogianni, W. Ketter, J. Collins, and D. Zhdanov, "Sustainable Electric Vehicle Charging using Adaptive Pricing," *Prod. Oper. Manag.*, vol. 29, no. 6, pp. 1550–1572, Jun. 2020.

- [20] M. Thalor and I. Abuj, "Smart Parking Management Assessment Using Machine Learning Algorithms and IOT." *International Journal of Scientific Research & Engineering Trends*, vol. 11, no.2, Apr 2025.
- [21] O. Bafail, "Optimizing Smart City Strategies: A Data-Driven Analysis Using Random Forest and Regression Analysis," *Applied Sciences (Switzerland)*, vol. 14, no. 23, Dec. 2024.
- [22] I. Mutambik, "Sustainable IoT-Enabled Parking Management: A Multiagent Simulation Framework for Smart Urban Mobility," *Sustainability (Switzerland)*, vol. 17, no. 14, Jul. 2025.
- [23] F. Piccialli, S. Amitrano, D. Cerciello, A. Borrelli, E. Prezioso, and M. Canzaniello, "A digital twin framework for urban parking management and mobility forecasting," *Nature Communications*, vol. 16, no. 1, Dec. 2025.
- [24] J. Chen, M. Wu, S. Li, Y. Cai, W. Long, and B. Yang, "Scale-Fusion Transformer: A Medium-to-Long-Term Forecasting Model for Parking Space Availability," *Electronics (Switzerland)*, vol. 14, no. 18, Sep. 2025.

Authors' Profiles



Siti Rhofiah is a final-year undergraduate student in Informatics Engineering at Telkom University, Purwokerto, Indonesia. She is currently completing her undergraduate thesis focusing on classification of parking revenue potential using Random Forest and spatial analysis. Her research interests include machine learning, data analytics, and spatial data analysis.



Sudioanto Sudioanto, he received a master's degree in computer science from IPB University, Indonesia. His research interests include Machine Learning and Remote Sensing. He is currently a Lecturer at the Department of Informatics, School of Computing, Telkom University, Purwokerto, Indonesia.



Aminatus Sa'adah is a lecturer at Telkom University, specializing in Applied Mathematics, focusing on mathematical modeling, optimization, and control theory. She completed her Master's degree in Mathematics at the Bandung Institute of Technology (2022) and earned her Bachelor's degree in Mathematics from Airlangga University (2018).

How to cite this paper:

Siti Rhofiah, Sudioanto Sudioanto, Aminatus Sa'adah, "A Random Forest Based Spatiotemporal Model for Parking Revenue Potential Classification to Support Adaptive Progressive Tariff Policies", *International Journal of Image, Graphics and Signal Processing(IJIGSP)*, Vol.18, No.4, pp. 106-120, 2026. DOI:10.5815/ijigsp.2026.04.06