STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
*Stat., Optim. Inf. Comput.*, Vol. **x**, Month 202x, pp 0–24.
Published online in International Academic Press (www.IAPress.org)

# Predicting Average Daily Rate for 4-5 Star Hotels and Resorts in Thailand: A Comparative Machine Learning Approach Integrating Physical Attributes and Site Characteristics

Anuwat Budda, Kongkoon Tochaiwat*

*Thammasat University, Pathum Thani, Thailand*

**Abstract** The rapid expansion of Online Travel Agencies (OTAs) has generated rich datasets that enable advanced analytical approaches to understanding hotel pricing behavior in tourism destinations. This study develops an interpretable machine learning framework to examine the determinants of hotel Average Daily Rate (ADR) in Thailand, a highly dynamic and spatially heterogeneous tourism market. A dataset comprising 500 four- and five-star hotels was compiled from Agoda and analyzed using five predictive models, including Multiple Linear Regression, Decision Tree, Gradient Boosting, Deep Learning, and Random Forest. Model performance was evaluated using 10-fold cross-validation. The Random Forest model achieved the highest predictive accuracy, with an $R^2$ of 0.606 and an RMSE of 40.99 USD. The results indicate that both linear and nonlinear modeling approaches contribute meaningfully to explaining hotel pricing structures, with ensemble-based methods demonstrating stable predictive performance in this context. Feature importance analysis reveals that Local Competitive Position, Star Rating, Room Size, Beach Location, and Facilities Total are the most influential determinants of ADR, reflecting the combined effects of intrinsic property attributes and contextual market positioning. To enhance interpretability and practical relevance, model outputs were further examined using local explanation techniques and validated through evaluation by experienced hospitality professionals. The expert assessment confirmed the practical coherence of the findings and their alignment with industry pricing logic. Overall, this study contributes to tourism and hospitality literature by integrating machine learning, explainable AI techniques, and expert validation to provide a context-sensitive framework for understanding hotel pricing dynamics in emerging tourism economies.

**Keywords** machine learning; Random Forest; hotel pricing; Average Daily Rate (ADR); Explainable AI; 4-5 star hotels

**DOI:** 10.19139/soic-2310-5070-4059

## 1. Introduction

Thailand comprises 76 provinces across six distinct regions, each characterized by diverse geographical features ranging from mountainous terrains to coastal and island destinations. Coastal and island regions such as Phuket and Krabi continue to serve as major attractions for international tourism [35], while mountainous provinces including Chiang Mai and Kanchanaburi have also experienced increasing popularity in recent years [50]. In response to evolving tourism dynamics, the Tourism Authority of Thailand (TAT) has implemented strategic initiatives aimed at promoting regional tourism development and supporting post-pandemic economic recovery. According to the Industry Outlook 2024–2026 published by Krungsri Research (2024), Thailand's hotel industry is expected to continue its recovery through 2026, with international tourist arrivals projected to return to approximately 38–40 million visitors by 2025, supported by ongoing government tourism promotion measures. These favorable market conditions have encouraged continued investment in the hospitality sector, leading both domestic and international hotel operators to expand into emerging tourism markets across Thailand.

*Correspondence to: Kongkoon Tochaiwat (Email: kongkoon@tu.ac.th).

ISSN 2310-5070 (online) ISSN 2311-004X (print)

However, this rapid development has led to an oversupply of accommodation, intensifying market competition. Hospitality investors face mounting challenges, including rising operational costs and economic uncertainty, which have caused numerous establishments to cease operations shortly after opening. In this hypercompetitive landscape, precise pricing strategies are no longer optional but essential for long-term viability.

While Artificial Intelligence (AI) and Machine Learning (ML) have been increasingly adopted in the hospitality industry for cancellation prediction [47] and reservation forecasting [14], similar analytical trends are observed in engineering and structural domains, including AI-driven construction management [52] and temperature-induced structural deformation modeling. Although Deep Learning has demonstrated immense capacity in modeling non-linear structural behaviors within complex engineering environments [16, 17], its empirical deployment on tabular, cross-sectional hospitality datasets warrants further comparative evaluation against tree-based ensemble methods. Consequently, integrating comparative frameworks with Explainable AI (XAI) tools like LIME serves as a robust approach to verify both predictive stability and local decision pathways in hotel pricing optimization.

Current research on hotel Average Daily Rate (ADR) has established a solid foundation using traditional regression models [51], which remain effective for processing structured data and providing transparent, interpretable results. These models offer valuable insights into linear relationships between price and hotel attributes. However, as the hospitality landscape becomes increasingly multifaceted, driven by complex interactions among location characteristics, site environments, and facility provision, the relationships between these variables often exhibit nonlinear patterns that conventional regression approaches may fail to fully capture.

Although room prices are readily observable through OTA platforms, published prices alone provide limited guidance for strategic planning because they do not explain why hotels with similar characteristics command different ADR levels or how individual hotel attributes contribute to pricing decisions. More importantly, published prices are unavailable for hotels in the planning, development, or renovation stages, where pricing decisions must be made before market data exist. Consequently, a predictive framework capable of estimating ADR from observable property characteristics offers practical value by supporting evidence-based decision-making during hotel development, design, market positioning, and revenue planning.

To address this empirical gap, this study deploys a comparative machine learning framework within the Altair AI Studio platform (Version 2026.0.4, Altair Engineering Inc., Troy, Michigan, USA), systematically evaluating five distinct modeling techniques: Multiple Linear Regression, Decision Tree, Gradient Boosting, Deep Learning, and Random Forest. This approach is intended not to replace traditional methodologies, but to extend the analytical capabilities of hospitality pricing research by identifying the exact algorithmic technique best suited to capturing the high-dimensional complexity of ADR determinants derived from OTA data. Furthermore, while previous studies have significantly contributed to the field by focusing on site-specific locations or niche geographic areas, this research adopts a broader national scope covering 4- and 5-star hotels across all major regions of Thailand, offering a comprehensive and nationally representative perspective on the structural factors driving ADR. To bridge the gap between computational outputs and industry reality, an expert validation phase is incorporated, synthesizing the best-performing model's predictive insights with the practical judgment of executive industry professionals to ensure that the identified pricing determinants are both statistically rigorous and managerially actionable. The proposed framework is intended to support evidence-based decision-making for multiple stakeholders, including hotel developers, architects, hotel operators, revenue managers, and tourism policymakers. By identifying the relative importance of hotel attributes influencing ADR, the findings can assist strategic planning, property development, pricing decisions, and tourism policy formulation.

## 2. Literature Review

### *2.1. Hotel Costs and Pricing Strategies*

The revenue structure of the hotel industry is primarily categorized into two distinct types. The core revenue is derived from room sales, which serves as the fundamental income source for hospitality businesses. In this context, the Average Daily Rate (ADR) is the standard metric used in the hotel industry to measure the average rental income per paid occupied room per day. Formally, ADR is defined as:

$$ADR = \frac{Room\ Revenue}{Rooms\ Sold} \tag{1}$$

where room revenue refers to the total income generated from occupied rooms, and rooms sold represents the number of paid occupied rooms within a given period. In the hospitality literature, ADR is widely used as a key performance indicator for pricing efficiency and revenue management [33]. Conversely, secondary revenue is generated from various amenities and services provided to both staying guests and external visitors, such as restaurants, spas, and entertainment venues like bars [9]

The relationship between hotel costs, specifically Average Daily Rates (ADR), and occupancy rates remains a cornerstone of effective hotel management. Research indicates that pricing strategies exert a significant influence on occupancy levels [19], highlighting the critical role of strategic marketing and price positioning in attracting customers and maximizing revenue within a competitive landscape.

To optimize occupancy rates, hotels must adopt flexible pricing strategies that account for shifting market demands and competitor pricing. However, a significant reduction in room rates may inadvertently lead to a decline in perceived value and overall profitability. Therefore, hotel management must strike a delicate balance between price, quality, and service standards to maintain brand integrity and guest satisfaction.

Beyond internal operational factors, external economic conditions also play a pivotal role in room rate determination. For instance, rising inflation directly impacts operating costs, while regions characterized by high purchasing power and intense demand often see a corresponding surge in ADR [55]. Consequently, geographic disparities are evident, with properties in rural areas often commanding lower price points compared to those in highly urbanized environments [49]

### *2.2. Determinants of Average Daily Rate (ADR)*

Table 1. Variables affecting ADR from the literature review.

| Author | Physical Attributes | Hotel Reputation | Location Attributes | Contextual Market Factors |
|---|---|---|---|---|
| [1] | | ✓ | | |
| [2] | | | ✓ | |
| [8] | | ✓ | ✓ | |
| [13] | | ✓ | ✓ | |
| [11] | ✓ | | ✓ | |
| [21] | ✓ | | | ✓ |
| [26] | ✓ | | | |
| [27] | | | ✓ | ✓ |
| [28] | ✓ | ✓ | | |
| [32] | | | ✓ | |
| [64] | ✓ | ✓ | | ✓ |
| [65] | ✓ | ✓ | | ✓ |
| [40] | | ✓ | | |
| [41] | ✓ | | | |
| [42] | ✓ | | ✓ | |
| [43] | ✓ | | | |
| [46] | | ✓ | | |
| [44] | ✓ | | | |
| [48] | ✓ | ✓ | | |
| [53] | | | ✓ | |
| [57] | ✓ | | ✓ | |
| [58] | ✓ | ✓ | ✓ | |
| [59] | | | ✓ | |
| [60] | ✓ | | ✓ | |
| [20] | | ✓ | | |
| [36] | ✓ | | | |
| [75] | | | | ✓ |
| [62] | ✓ | | | |
| [71] | ✓ | | | ✓ |
| [61] | ✓ | | ✓ | |
| [18] | ✓ | | | |
| [74] | ✓ | ✓ | ✓ | ✓ |
| [7] | ✓ | | | |
| [3] | ✓ | | | |
| [10] | ✓ | | | |

The determination of ADR is a multi-dimensional process influenced by various factors that guests evaluate when perceiving value. In this research, these determinants are categorized into four major groups: Physical Attributes, Hotel Reputation, Location Attributes, and Market Positioning. To provide a comprehensive overview of how these dimensions have been globally operationalized, a structural synthesis of foundational pricing literature is systematically illustrated in Table 1, mapping the research focus of prior scholars across these four macro dimensions.

*2.2.1. Physical Attributes* Physical attributes refer to the tangible structural components of a hotel property that guests perceive and experience post-arrival, heavily impacting both post-stay customer satisfaction and foundational property construction costs. Extant literature establishes that variations in hotel pricing frequently reflect localized property types and structural configurations, as distinct architectural layouts require highly specialized design paradigms and varying capital investments ([36]). As extensively synthesized in Table 2, historical hedonic pricing studies have routinely isolated individual, discrete physical amenities such as swimming pools, fitness centers, spa facilities, and specialized conference spaces as independent drivers of room rate appreciation.

Table 2. Hotel physical attributes influencing ADR from literature reviews.

| **Authors** | **Physical Attributes Variable** |
|---|---|
| [4]; [45] | Spa |
| [4]; [41]; [43] | Restaurant |
| [4]; [45]; [43] | Gym |
| [3]; [10]; [26]; [61]; [44] | Room Quality |
| [41]; [43] | Parking |
| [41]; [43] | Conference |
| [43] | Sauna |
| [43]; [10]; [26]; [61] | Bar and Lounge |
| [61] | Pool |
| [37] | Number of rooms |

However, to bypass the mathematical constraints of high-dimensional multicollinearity and to optimize algorithmic processing within the modern machine learning environment, this study innovatively consolidates these separate physical amenities into a single aggregated index variable quantified as Facilities Total. This global index accurately captures the architectural and service density of the property. Concurrently, consistent with the foundational frameworks of [3] and [10], physical room properties are represented via Room Size. Rather than viewing floor area strictly through dimensional metrics, this study aligns with literature positing that room dimensions operate as an institutional proxy for holistic room quality, encompassing environ-mental parameters, acoustic insulation, and spatial functionality [62].

*2.2.2. Hotel Reputation* Hotel reputation refers to the brand image that fosters consumer trust, safety, and expectations, all of which directly influence ADR. A higher Star Rating signals a commitment to superior standards, leading guests to expect premium quality in room conditions, aesthetic appeal, and service excellence. Consequently, guests are willing to pay a higher premium for these guaranteed standards [20]. Furthermore, Chain Affiliation plays a crucial role in ADR through stringent ”Design Standards” that meet international requirements. Managed brands often re-quire higher initial investments in design and construction compared to independent hotels to maintain brand identity and cater to specific global target segments. The criteria for hotel star classification are derived from multiple established sources. According to studies by [30] and [38], hotel classification systems worldwide are predominantly governed by official hotel rating systems. These systems, often operated by government agencies or authorized national organizations, employ clearly disclosed evaluation criteria that are accessible to the public. The implementation of such official ratings ensures that a higher star designation inherently reflects a superior and more rigorous hotel standard. Beyond serving as a vital information source for consumers, these official classifications provide a formalized framework for categorizing hotels, which also serves administrative purposes such as tax rate determination and industry benchmarking.

*2.2.3. Location Attributes* Location attributes encompass the geographical characteristics of a hotel's site, which provide both ad-vantages and constraints. These factors significantly impact the guest's travel experience and the hotel's competitive economic standing. For instance, Urban Locations offer convenience and proximity to shopping districts and metropolitan views [46]. In contrast, Beachfront Locations provide exclusive environmental value, such as ocean views and private beach access ([53]). These spatial advantages allow hotels to exercise pricing power based on exclusivity and accessibility, which are key drivers of ADR.

*2.2.4. Local Competitive Position* During the hotel selection process, travelers typically evaluate alternative accommodations by simultaneously considering location, facilities, service quality, and price. Consequently, guests' willingness to pay reflects the over-all value of a hotel relative to competing properties within the same market rather than being determined by any single attribute alone [27]. This suggests that hotel pricing should be understood not only through intrinsic property characteristics but also within the broader competitive context in which the hotel operates.

Previous studies have shown that hotel pricing is strongly influenced by competitive positioning. Rather than determining room rates solely on the basis of construction cost or service quality, hotels frequently adopt competitor-based pricing strategies by positioning their prices relative to comparable properties operating within the same market. [75] demonstrated that hotels maintaining room rates above those of their direct competitors achieved higher revenue performance, highlighting the strategic importance of relative market positioning in hotel pricing decisions. More recently, [74] proposed a data-driven competitor profiling framework using multidimensional OTA information, emphasizing that competitive relationships among hotels can be effectively represented through external market information rather than relying solely on individual hotel characteristics.

Building upon these studies, the present research introduces Local Competitive Position (LCP) as a contextual market factor describing a hotel's relative pricing position within its provincial competitive environment. The variable was constructed using independently collected OTA market information from comparable four- and five-star hotels located within the same province. Hotels with published room rates above the provincial market median were classified as having a higher local competitive position. Unlike traditional hotel attributes that describe the physical or operational characteristics of an individual property, Local Competitive Position represents the competitive context surrounding each hotel and complements architectural, locational, and reputational attributes by incorporating externally observed market information into the prediction framework.

### 2.3. Machine Learning Techniques for Hotel ADR Prediction

*2.3.1. Multiple Linear Regression* Multiple Linear Regression (MLR) is a foundational statistical technique that models the linear relationship between a dependent variable and two or more independent variables. [69] In the context of hotel pricing research, MLR has been widely applied as a baseline model to examine how property attributes and location characteristics influence ADR, owing to its interpretability and computational efficiency. Linear regression, as a foundational statistical model, contains various extensions for flexible and strong modeling and has been consistently applied in hospitality forecasting research as a benchmark against which more complex models are evaluated. Despite its simplicity, MLR provides a useful reference point for assessing the marginal contribution of non-linear modeling approaches.

*2.3.2. Decision Tree* A Decision Tree is a non-parametric supervised learning algorithm that partitions the feature space into hierarchical regions through a series of binary splits, generating rule-based predictions that are highly interpretable. A key strength of decision trees lies in their exceptionally interpretable, intuitive tree-like structure that enhances understanding and communicability of results [70]. While individual decision trees are prone to overfitting on small datasets, they serve as a transparent and intuitive baseline for nonlinear modeling and form the building block of more advanced ensemble methods, including the various resampling-based multiple-tree approaches that have emerged from the cross-fertilization between statistical and machine learning communities [70, 68].

*2.3.3. Gradient Boosting* Gradient Boosting is a sequential ensemble technique originally formulated by [67] through a gradient descent framework in function space, in which a general boosting paradigm is developed for additive expansions based on any fitting criterion. Rather than optimizing model parameters directly, the algorithm constructs an additive expansion of weak learners, typically shallow decision trees, in which each successive tree is trained to correct the residual errors of its predecessor. The modeling process proceeds iteratively: an initial tree is first fitted to the target variable, and in each subsequent iteration, a new tree is grown to approximate the negative gradient of the loss function with respect to the current ensemble prediction. The final prediction for each observation is ex-pressed as the weighted sum of outputs across all trees in the ensemble, gradually forming a strong predictor through the accumulation of many weak models. [67] demonstrated that this approach produces competitive, highly robust, and interpretable procedures for both regression and classification tasks and is particularly appropriate for mining less-than-clean data. Building upon this foundation, [73] developed XGBoost, a scalable implementation of gradient-boosted trees that extends the original algorithm through regularization techniques and computational efficiency improvements, achieving state-of-the-art performance in classification and regression tasks across a wide range of structured datasets. Gradient-boosted trees have been increasingly applied in price prediction tasks as an alternative to hedonic regression models, demonstrating competitive predictive accuracy on structured property attribute data. Its ability to capture nonlinear relationships and feature interactions makes it particularly well-suited for hotel pricing datasets characterized by mixed numerical and categorical variables.

*2.3.4. Deep Learning* Deep Learning (DL) represents a class of artificial neural network architectures comprising multiple hidden layers, enabling the automated extraction of increasingly abstract feature representations from raw input matrices. While standard Artificial Neural Networks (ANN) provide the foundation for biologically inspired computing [39], the increased depth in DL topologies allows for high accuracy in complex forecasting tasks by identifying latent relationships between diverse input variables and the target output. Recent hospitality-specific studies have validated its computational capacity in complex applications, including the analysis of guest satisfaction drivers and online review patterns [48, 54]. Within the context of ADR prediction, where pricing structures are influenced by a multitude of intersecting factors ranging from physical amenities to macro-market positioning dynamics, the multilayered architecture of DL presents a robust analytical capacity to decode complex, non-linear pricing relationships alongside ensemble methods.

*2.3.5. Random Forest* Random Forest is a bagging ensemble method introduced by [66] that constructs a large number of decision trees on random subsets of the training data and aggregates their predictions through averaging, thereby reducing variance and improving generalization. Random Forest has been demonstrated to outperform conventional regression approaches in price prediction tasks, with its capacity to handle nonlinear relationships and high-dimensional data contributing to strong predictive performance across diverse property datasets. In the hospitality domain, Random Forest has been applied to hotel-related prediction tasks, demonstrating its effectiveness in capturing the complexity of accommodation pricing data. The model's built-in feature importance mechanism additionally provides interpretable insights into the relative contribution of each variable, supporting both prediction and explanation objectives.

## 3. Methodology

### *3.1. Research Design*

This study employs a quantitative research design to develop and evaluate predictive models for hotel Average Daily Rate (ADR). A systematic literature review was first conducted to establish a theoretical foundation and identify variables with documented influence on hotel pricing. Candidate variables were subsequently refined based on the availability of publicly accessible data, ensuring the practical applicability of the proposed framework.

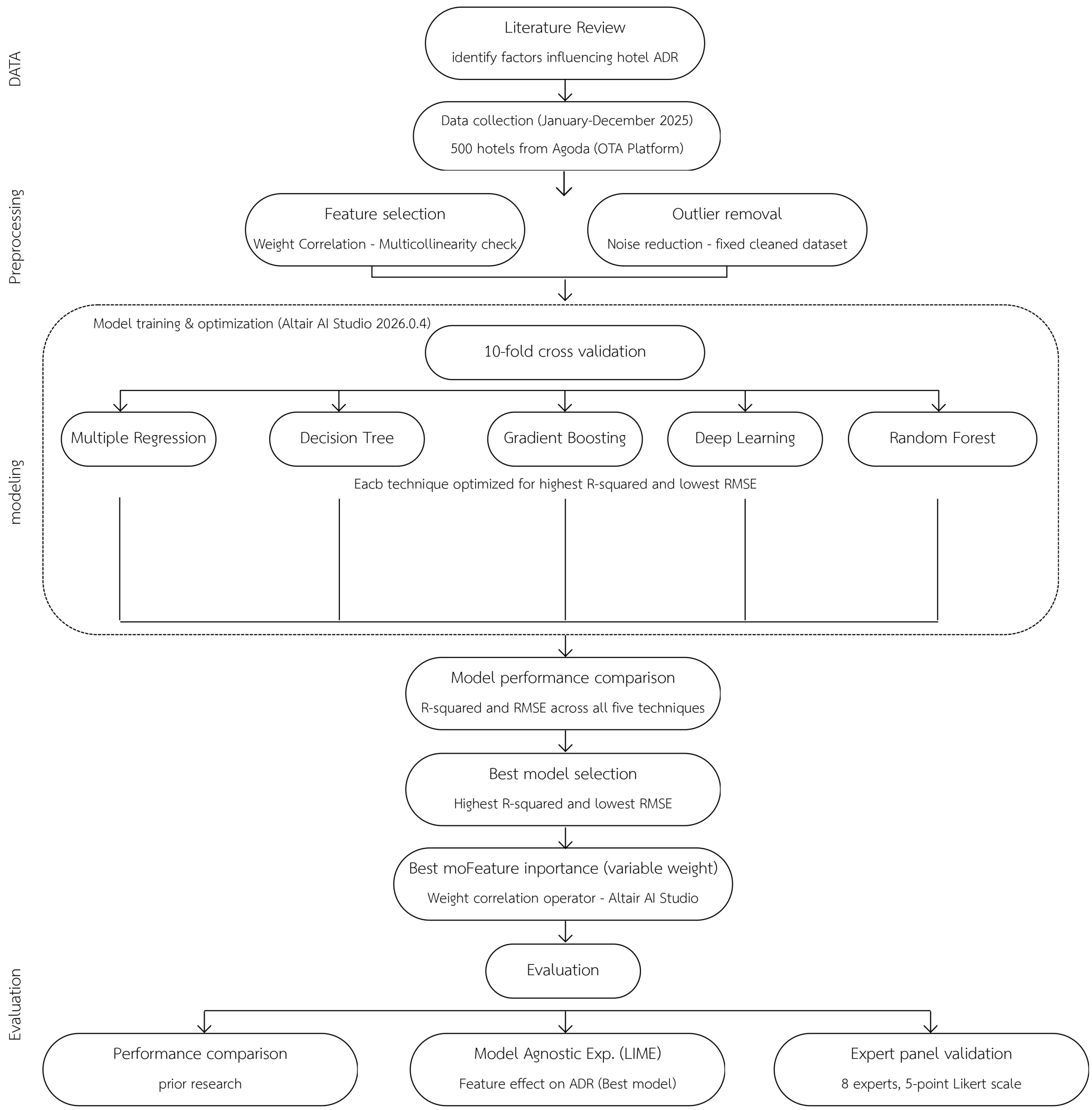


Figure 1. Research Framework diagram

### 3.2. Data Collection

A dataset comprising 500 four- and five-star hotels and resorts was constructed by extracting property-level information from Agoda, one of the world's largest Online Travel Agency (OTA) platforms. The dataset provided nationwide coverage across Thailand and included properties located in four site characteristic categories: City, Beach, Nature, and Other/General locations, with the latter serving as the reference category. The selected variables were classified into two groups. Numerical variables included explicitly quantifiable attributes such as the total number of guestrooms, average room size, total facilities, and distance from the city center. Categorical variables were obtained from Agoda's standardized property classification system and included hotel type, star rating, chain affiliation, site characteristics, and location-related attributes. Room price data were collected directly from the Agoda platform between January and December 2025. To minimize the influence of short-term price fluctuations associated with promotional campaigns, seasonal demand, and temporary market conditions, the published room rate of each property was recorded three times per month (on the 1st, 15th, and 30th), resulting in repeated observations over a 12 months period. The arithmetic mean of all recorded published room rates was subsequently calculated and used as the representative Average Daily Rate (ADR) for each property. A structured selection procedure was applied to construct the final dataset of 500 hotels based on filtered listings from the Agoda Online Travel Agency (OTA) platform for four- and five-star hotels in Thailand. A consistent filtering and ordering process was used across all observations. Based on this ordered list, every 10th hotel was selected after a randomly determined starting point between 1 and 10. This procedure provides a systematic and evenly distributed coverage of hotel listings under consistent extraction conditions, without assuming strict probabilistic sampling. To capture differences in hotel pricing across geographically diverse markets, an additional contextual variable termed Local Competitive Position (LCP) was constructed. For each province, the median published room rate of comparable four- and five-star hotels was independently estimated using additional OTA market data. Hotels with published room rates above the provincial median were classified as having a high Local Competitive Position, whereas those at or below the provincial median were classified as having a low Local Competitive Position. The LCP variable serves as a local market benchmark, reflecting whether a hotel's published room rate is positioned above or below the pre-vailing market price within its provincial market. The variable was introduced to account for the spatial heterogeneity of Thailand's hotel market. Because hotel pricing is influenced by local economic conditions, tourism demand, land values, and competitive intensity, similar room rates may represent different market positions across provinces. Benchmarking each hotel against comparable properties within the same province therefore provides contextual market information that enables the prediction model to distinguish hotels operating under different local market conditions.

### 3.3. Data Preprocessing

Before model training, the Weight Correlation operator in Altair AI Studio was applied to assess the linear relationship between each independent variable and ADR. Variables exhibiting low correlation weight or high inter-variable correlation indicating redundancy or multicollinearity were excluded from the feature set, ensuring that only variables with meaningful predictive contribution and acceptable inter-variable independence were retained for model training.

### 3.4. Model Training and Optimization

Five modeling approaches were trained and evaluated using Altair AI Studio (Version 2026.0.4, Altair Engineering Inc., Troy, Michigan, USA), a platform selected for its integrated visual modeling environment, enabling systematic comparison of multiple machine learning techniques within a unified and reproducible analytical workflow. These techniques were selected to represent a spectrum of modeling paradigms, from linear and rule-based approaches to sequential ensemble and deep neural architectures, enabling a comprehensive comparative assessment of predictive accuracy and interpretability. A 10-fold cross-validation procedure was embedded within the training of each technique to minimize overfitting and ensure reliable performance estimation. Each technique was individually optimized with the objective of maximizing the coefficient of determination ($R^2$) and minimizing Root Mean Square Error (RMSE). The five techniques are as follows.

Multiple Regression serves as the linear baseline, providing a benchmark against which more complex non-linear models are evaluated. The model incorporated the M5 prime feature selection method to identify the most significant predictors. To ensure the integrity of the regression coefficients, the 'eliminate collinear features' option was enabled with a minimum tolerance of 0.05 to mitigate multicollinearity issues. The model included a bias term (intercept) to accurately represent the baseline value of the dependent variable.

Decision Tree employs a rule-based, nonlinear structure capable of capturing hierarchical feature interactions in hotel pricing data. To improve the predictive performance of the Decision Tree model, the splitting criterion was transitioned from Gain Ratio to Least Squares. This adjustment allows the algorithm to minimize the sum of squared differences between the predicted and actual values, which is more effective for regression tasks. Additionally, the maximal depth of the tree was increased to 15 to enable the model to capture more complex interactions within the dataset. To balance the learning capacity, the pre-pruning constraints were relaxed by setting a lower minimal gain and reduced minimal leaf size, allowing the tree to grow sufficiently to represent the underlying patterns in the Hotel ADR.

Random Forest: The study employed Random Forest as a bagging ensemble method to capture nonlinear relationships between the independent variables and hotel ADR. The model configuration in Altair AI Studio utilized 200 trees to ensure stable and reliable predictions through aggregation of a large number of decision trees. The splitting criterion was set to least squares, which is the standard criterion for regression tasks, as it minimizes the sum of squared residuals at each node. The maximal tree depth was set to 20 to allow sufficient complexity for each individual tree to capture higher-order feature interactions. The subset ratio for feature sampling at each split was determined automatically using the guess subset ratio function, which follows the conventional heuristic of selecting the square root of the total number of features, thereby ensuring diversity among individual trees and reducing inter-tree correlation. The voting strategy was set to average vote, whereby the final ADR prediction is computed as the arithmetic mean of predictions from all 200 trees, which is the appropriate aggregation method for continuous regression targets.

Gradient Boosting extends the Decision Tree framework through sequential ensemble learning, in which each successive tree corrects the residual errors of its predecessor, yielding strong predictive performance on tabular data. To capture potential nonlinear relationships and enhance predictive accuracy, a Gradient Boosted Trees (GBT) algorithm was implemented. The GBT model was configured with an ensemble of 50 trees and a learning rate of 0.01. To prevent overfitting, an early stopping strategy was applied, which halts the training process once the model's performance on the validation set stabilizes. Furthermore, to ensure the reproducibility of the results, a fixed local random seed was utilized, maintaining consistency across all computational runs. The Gaussian distribution was selected for the loss function to optimize the regression performance for continuous data.

A deep learning model based on a feedforward neural network architecture was employed to capture high-order nonlinear relationships between the independent variables and hotel ADR. The network comprised three hidden layers with 128, 64, and 32 neurons, respectively, progressively reducing in size to extract increasingly abstract feature representations from the input data. The Rectifier (ReLU) activation function was applied across all hidden layers to mitigate the vanishing gradient problem and accelerate convergence. The output layer employed a linear activation function, consistent with the continuous regression target of ADR prediction. The model was trained over 200 epochs with an adaptive learning rate enabled, allowing the optimization process to automatically adjust the step size based on gradient behavior. The reproducible option was enabled to ensure consistent results across repeated runs by fixing the random seed.

### *3.5. Performance Evaluation Metrics*

Upon completion of optimization, the predictive performance of all five techniques was systematically compared based on cross-validated $R^2$ and RMSE. The technique yielding the highest $R^2$ and lowest RMSE was designated as the best-performing model and selected for subsequent interpretability analysis. Model performance was assessed using two standard evaluation metrics: the coefficient of determination ($R^2$) and Root Mean Square Error (RMSE). $R^2$ quantifies the proportion of variance in the target variable (ADR) that is explained by the model, with values closer to 1.0 indicating superior predictive accuracy. RMSE measures the average magnitude of prediction errors in the original unit of the target variable, with lower values indicating better model fit. These metrics are calculated as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_{i,o} - y_{i,p})^2} \tag{2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_{i,o} - y_{i,p})^2}{\sum_{i=1}^{n}(y_{i,o} - \bar{y}_o)^2} \tag{3}$$

Where n is the total number of observations, y_(i,o) is the observed ADR value of the i-th property, y_(i,p) is the predicted ADR value of the i-th property, and $\bar{y}_o$ is the mean of all observed ADR values. All model evaluations were conducted using 10-fold cross-validation to ensure reliable and generalizable performance estimation.

### 3.6. Feature Importance Analysis

The relative contribution of each input variable to ADR prediction was quantified using the Weight Correlation operator applied to the best-performing model in Altair AI Studio. This analysis identifies the variables exerting the greatest influence on hotel pricing and provides an empirical basis for managerial interpretation.

### 3.7. Model Evaluation

Model evaluation was conducted through three complementary approaches. First, the predictive performance of the best model was benchmarked against results reported in prior studies to contextualize its accuracy within the existing literature. Second, Local Interpretable Model-Agnostic Explanations (LIME) were generated using the Model Simulator operator in Altair AI Studio, applied exclusively to the best-performing model. This analysis examined the directional effect and magnitude of each feature on individual ADR predictions, providing local interpretability beyond aggregate importance scores. As a model-agnostic method, LIME is applicable regardless of the internal structure of the selected model [63]. Third, an expert panel validation was conducted to assess the practical credibility of the feature importance findings, as described in Section 3.8.

### 3.8. Expert Panel Validation

To validate the practical relevance of the feature importance results, a structured expert panel was conducted with eight hotel industry professionals. Panelists were selected according to two criteria: a minimum of 10 years of professional experience within the upscale hotel sector, and a current tenure in a senior leadership, general management, or director-level role specializing in operations, branding, or revenue and commercial management. Each expert independently evaluated the feature importance findings using a structured questionnaire administered on a five-point Likert scale (1 = strongly disagree, 5 = strongly agree). Mean scores were calculated for each item, with a threshold of $\geq$ 3.50 adopted as the criterion for acceptable validity and practical applicability.

### 3.9. Expert Validation

To ensure the reliability, practical efficacy, and professional credibility of the developed predictive framework, a panel of eight industry experts was selected to evaluate the feature importance findings and their practical applications. The panel was assembled based on rigorous purposive sampling criteria to guarantee high-level strategic and operational insights. Each expert was required to meet two strict inclusion criteria: (1) a minimum of 10 years of professional experience within the upscale (4-to-5 star) hotel sector, and (2) a current tenure in a senior leadership, general management, or director-level role specializing in hotel operations, branding, or revenue and commercial management. The detailed professional profiles of the selected experts are summarized in Table 3.

Table 3. Expert Professional Profiles.

| Expert No. | Position | Years of Experience | Organization Type / Expertise Context |
|---|---|---|---|
| 1 | Advisor to the Board of Directors | 15 | Hotel Management Service Company |
| 2 | CEO | 15 | Hotel Management Service Company |
| 3 | CEO | 12 | Hotel Branding Company |
| 4 | Division Manager | 20 | Integrated Hotel Business Consultancy |
| 5 | Sales Director | 25 | Luxury Hotel, Resort, and Spa Operations |
| 6 | General Manager | 20 | International 5-Star Hotel Chain |
| 7 | General Manager | 20 | International 5-star Hotel Chain |
| 8 | General Manager | 20 | International 5-star Hotel Chain |

## 4. Result

### *4.1. Relevant Variables*

Table 4. Summary statistics

| Variable Group | Variable name | Variable Description | Measurement Scale | Unit |
|---|---|---|---|---|
| Hotel Reputation | Chain Hotel | Status as a chain-managed hotel | Dummy Variable | Yes=1/No=0 |
| | Star (4 or 5) | Hotel star rating | Ordinal | 4 or 5 stars |
| Contextual Market Factors | Local Competitive Position | Binary variable indicating whether a hotel's published room rate exceeds the median room rate of comparable 4–5-star hotels within the same province, based on independently collected OTA market data. | Dummy Variable | Yes=1/No=0 |
| Location Attributes | Site Character Beach | Physical location characterized by beachfront | Dummy Variable | Yes=1/No=0 |
| | Site Character City | Physical location characterized by urban/city setting | Dummy Variable | Yes=1/No=0 |
| | Site Character Nature | Physical location characterized by natural surroundings | Dummy Variable | Yes=1/No=0 |
| | Travel City | Located in a designated tourist city | Dummy Variable | Yes=1/No=0 |
| | Travel Spot nearby | Presence of tourist attractions within a 1 km radius | Dummy Variable | Yes=1/No=0 |
| | Distance from City Center in 1km. | Distance from the city center (within 1 km range) | Dummy Variable | Yes=1/No=0 |
| Physical Attributes | Hotel type Hotel | Functional classification as a Hotel | Dummy Variable | Yes=1/No=0 |
| | Hotel type Resort | Functional classification as a Resort | Dummy Variable | Yes=1/No=0 |
| | Facility Total | Total number of hotel facility attributes explicitly reported in OTA listings, (e.g., spa services, fitness centers, swimming pools, lounges, entertainment facilities, conference rooms, restaurants, and bars). | Numerical (Integer) | Count |
| | Room Size | Total floor area of the guest room | Numerical (Continuous) | Square Meters ($m^2$) |
| | Room Total | Total number of guest rooms | Numerical (Integer) | Count |

Based on the synthesis of variables identified in Section 2, all candidate predictors were initially integrated into Altair AI Studio (Version 2026.0.4, Altair Engineering Inc., Troy, Michigan, USA) for preliminary analysis. These variables were categorized into four primary groups: Location, Hotel Reputation, Contextual Market Factors, and Physical Attributes. Preliminary feature screening was conducted using the Weight by Correlation operator, which assessed the linear relationship between each independent variable and the target variable (ADR). Variables exhibiting low predictive correlation or high inter-variable correlation indicative of redundancy or multicollinearity were excluded from the final feature set. Multicollinearity refers to a condition in which two or more independent variables are highly correlated, which may inflate model complexity and reduce the stability of coefficient estimates. This screening procedure is functionally equivalent to Pearson correlation-based feature screening combined with Variance Inflation Factor (VIF) analysis, ensuring that only variables with meaningful and independent predictive contribution were retained. The finalized set of input features and their respective formats are detailed in Table 4.

Based on the variables presented in Table 4, the data can be categorized into four fundamental groups that collectively influence hotel pricing structures.

1) Hotel Reputation comprises the ”Chain Hotel” and ”Star Rating” variables. The inclusion of the chain hotel variable is significant as it reflects the operational credibility and standardized service quality inherent in managed brands. These hotels are expected to meet rigorous design and construction standards, thereby aligning the physical property with the high expectations of brand-conscious guests. Similarly, the 4-star and 5-star ratings serve as indicators of premium quality, signaling an increased variety and sophistication of facilities that justify higher price points.

2) Contextual Market Factors comprise the Local Competitive Position variable. This variable indicates whether a hotel's published room rate exceeds the median room rate of comparable 4–5-star hotels within the same province, based on independently collected OTA market data. Rather than representing the hotel's intrinsic characteristics, this variable captures its relative pricing position within the local competitive environment. Given the substantial regional variation in tourism demand, land values, and economic conditions across Thailand, hotels with similar physical characteristics may adopt different pricing levels depending on their local market context. ([74]; Enz et al., 2016) Therefore, this variable provides complementary contextual information beyond hotel-specific attributes.

3) Location Attributes play a pivotal role in determining a hotel's market value. This group encompasses variables reflecting the physical environment and geographic setting of each property, including site characteristics such as beach, city, and nature surroundings, proximity to tourist attractions within a 1 km radius, location within a designated travel city, and distance from the city center. These spatial factors directly reflect the underlying cost of land, which fluctuates significantly across different geographical regions. Consequently, variation in land value acts as a primary driver of a hotel's base pricing strategy, as properties in high-demand areas must account for higher investment and operational costs.

4) Physical Attributes represent the tangible characteristics of the hotel, including Room Total, Room Size, and Facilities Total. The Facilities Total variable represents the total number of hotel facility attributes available to guests, constructed by aggregating all facility-related attributes explicitly reported in the Agoda OTA listings. Each observable facility item (e.g., spa services, fitness center, swimming pools, restaurants, and conference facilities) is counted as a unit, and multiple instances of the same facility type are included when individually reported in the listing. Collectively, these variables describe the physical capacity and service infrastructure of the hotel, which may contribute to differences in perceived value and room pricing.

### *4.2. Descriptive statistics dataset*

To understand the dataset better, Tables 5 and 6 present the descriptive statistics for all variables included in this study.

The dataset comprises 500 hotel and resort properties collected from the Agoda OTA platform. The mean ADR was 101.87 USD with a standard deviation of 72 USD, reflecting substantial price variation across properties in the sample. On average, properties offered 5.53 total facilities, contained 134.67 rooms, and provided a mean room size of 35.83 $m^2$, with considerable variability in room total (SD = 125.79) and room size (SD = 24.22) indicating a diverse range of property scales.

Table 5. Descriptive statistics of study continuous variables (n=500)

| Variable | n | Mean | SD | Min | Max |
|---|---|---|---|---|---|
| ADR (USD) | 500 | 101.87 | 72 | 27 | 276 |
| Facilities total | 500 | 5.53 | 2.76 | 0 | 17 |
| Room total | 500 | 134.67 | 125.79 | 10 | 760 |
| Room size ($m^2$) | 500 | 35.83 | 24.22 | 14 | 300 |

Table 6. Descriptive statistics of study categorical variables (n=500)

| Variable | Category | n | % |
|---|---|---|---|
| Chain hotel | Yes (1) | 162 | 32.4 |
| | No (0) | 338 | 67.6 |
| Star rating | 5 stars | 164 | 32.8 |
| | 4 stars | 336 | 67.2 |
| Hotel type | Hotel | 323 | 64.6 |
| | Resort | 177 | 35.4 |
| Site Character | Other/General | 185 | 37.0 |
| | City | 157 | 31.4 |
| | Beach | 91 | 18.2 |
| | Nature | 67 | 13.4 |
| Travel city | Yes (1) | 410 | 82.0 |
| | No (0) | 90 | 18.0 |
| Travel spot nearby | Yes (1) | 362 | 72.4 |
| | No (0) | 138 | 27.6 |
| Distance from city center | Yes (1) | 149 | 29.8 |
| | No (0) | 351 | 70.2 |
| Local Competitive Position | Yes (1) | 274 | 54.8 |
| | No (0) | 226 | 45.2 |
| Region | Central | 135 | 27 |
| | Northern | 107 | 21.4 |
| | Northeastern | 32 | 6.4 |
| | Eastern | 26 | 5.2 |
| | Western | 28 | 5.6 |
| | Southern | 172 | 34.4 |

Regarding categorical characteristics, the majority of properties were independent hotels (67.6%) rather than chain-affiliated (32.4%), and most were classified as hotels (64.6%) rather than resorts (35.4%). Four-star properties constituted the largest proportion (67.2%), with five-star properties comprising the remaining 32.8%. In terms of location, the largest site character group was Other/General locations (37.0%), followed by City (31.4%), Beach (18.2%), and Nature (13.4%). Most properties were situated within designated travel cities (82.0%) and in proximity to tourist attractions (72.4%), while 29.8% were located within 1 km of the city center. With respect to market positioning, 54.8% of properties were classified as having a high Local Competitive Position (LCP), while 45.2% were classified as low LCP. The regional distribution shows a higher concentration of observations in the Central (27.0%) and Southern (34.4%) regions, followed by the Northern region (21.4%). The regional distribution also indicates that the dataset covers all major regions of Thailand, although a larger proportion of observations originates from the Central and Southern regions, reflecting the higher concentration of four- and five-star hotel developments in these major tourism and economic destinations.

### 4.3. Model Performance Comparison

Table 7. Comparison of Model

| Metric | Random forest | Multiple Linear Regression | Gradient Boosting | Decision Tree | Deep Learning |
|---|---|---|---|---|---|
| $R^2$ | 0.606 | 0.575 | 0.559 | 0.547 | 0.473 |
| RMSE | 40.99 USD | 42.67 USD | 45.92 USD | 44.43 USD | 50.12 USD |

The predictive performance of the five optimized machine learning models was evaluated using 10-fold cross-validation, and the results are summarized in Table 7. Among the evaluated algorithms, Random Forest (RF) achieved the highest predictive performance, yielding the highest coefficient of determination ($R^2$ = 0.606) and the lowest prediction error (RMSE = 40.99 USD). Multiple Linear Regression (MLR) ranked second ($R^2$ = 0.575, RMSE = 42.67 USD), followed by Gradient Boosting (GBT) ($R^2$ = 0.559, RMSE = 45.92 USD), Decision Tree (DT) ($R^2$ = 0.547, RMSE = 44.43 USD), and Deep Learning (DL) ($R^2$ = 0.473, RMSE = 50.12 USD).

The predictive performance obtained in this study is comparable to previous hotel pricing studies employing cross-sectional hotel characteristics. For example, [76] reported an adjusted $R^2$ of 0.537 using a log-linear hedonic pricing model, while [79] achieved an $R^2$ of 0.460 using a Gradient Boosting Machine for hotel room price prediction. The Random Forest model developed in the present study achieved a higher explanatory capability ($R^2$ = 0.606) than these studies. Although [65] reported a higher $R^2$ of 0.806 using Random Forest, their model incorporated a different dataset and variable structure. Likewise, [80] demonstrated that predictive performance varied substantially across hotel market contexts, with $R^2$ values ranging from 0.350 for resort hotels to 0.861 for city hotels. These findings indicate that predictive accuracy is strongly influenced by differences in dataset characteristics and model specifications across studies.

### 4.4. Feature Importance Analysis

To evaluate the relative importance of each predictor in estimating hotel Average Daily Rate (ADR), feature importance analysis was performed using the Weight by Correlation operator in Altair AI Studio. This approach estimates the strength of the association between each predictor and the target variable, assigning a normalized weight that reflects its relative contribution to the prediction task. The analysis included both hotel-specific attributes (e.g., star rating, room size, facilities, and architectural characteristics) and contextual market variables, including Local Competitive Position. The normalized feature importance values and their corresponding rankings are presented in Figure 2.

Higher normalized weights indicate greater overall importance in the prediction model; however, these values should be interpreted as measures of relative predictive contribution rather than causal effects. The feature im-portance analysis identified Local Competitive Position as the most influential predictor, with the highest normalized weight (1.000). This finding indicates that a hotel's pricing position relative to the prevailing local market benchmark provides the most important contextual information for predicting Average Daily Rate (ADR). Rather than reflecting the absolute room price, this variable benchmark each hotel against comparable four- and five-star properties within the same provincial market, enabling the model to account for regional market heterogeneity. Consequently, hotels with similar physical characteristics may exhibit different ADRs because they are positioned relative to different local market benchmarks. Among the hotel-specific attributes, Star Rating (0.796) and Room Size (0.741) were the most influential predictors, followed by Site Character: Beach (0.552) and Facilities Total (0.420). These results indicate that both official hotel classification and the physical scale of accommodation con-tribute substantially to ADR prediction, while environmental setting and the overall intensity of facility provision provide additional predictive information. In this study, Facilities Total reflects the aggregated number of facility attributes explicitly reported in OTA listings, capturing the extent of amenity availability and service capacity within each property. Hotels with a higher level of facility provision are likely to offer broader service coverage and enhanced guest experience, which may translate into stronger pricing power in the market. Conversely, Site Char-acter: City (0.028) and Located in City Center (0.000) exhibited relatively low global

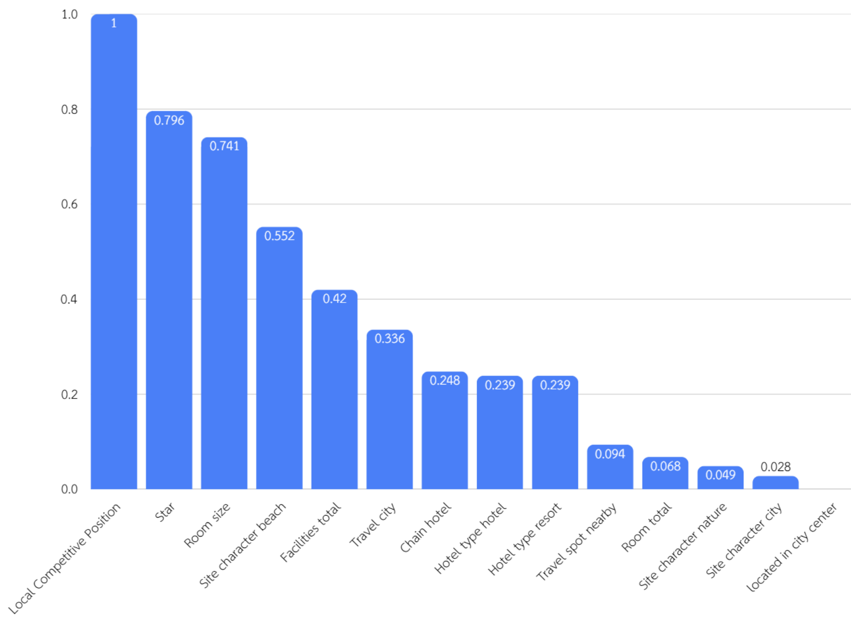


Figure 2. Normalized feature importance derived from the Random Forest model, illustrating the relative importance of hotel attributes and contextual market variables in predicting Average Daily Rate (ADR).

importance. This finding does not necessarily imply that these variables lack relevance. Instead, their predictive information may already be captured by other correlated variables within the model, particularly Local Competitive Position, which incorporates provincial-level market benchmarking. As a result, the marginal contribution of these location-related variables is reduced in the Random Forest model due to overlapping explanatory information. To address concerns regarding model dependency in feature importance estimation, a cross-model comparison was conducted across all implemented machine learning algorithms, including Multiple Linear Regression, Decision Tree, Random Forest, Gradient Boosting, and Deep Neural Network models. The results demonstrate a highly consistent ranking structure across all models. Specifically, Local Competitive Position, Star Rating, and Room Size consistently appear as the top three most important predictors, while lower-ranked variables show minor variations in relative ordering. This strong consistency indicates that the identified feature importance structure is robust and not dependent on a single modeling approach, but rather reflects stable underlying relationships within the dataset.

Overall, the feature importance analysis demonstrates that hotel pricing is jointly associated with two complementary dimensions. The first comprises intrinsic hotel characteristics, including star rating, room size, and facility total, whereas the second is represented by Local Competitive Position, which provides a contextual market benchmark that enables the model to distinguish hotels operating under different regional market conditions across Thailand.

### *4.5. LIME-Based Model Interpretability Analysis*

To improve the interpretability of the Random Forest model, a local explanation analysis was performed using the Model Simulator operator. This approach approximates the behavior of the predictive model around an individual observation, allowing the contribution of each feature to the predicted Average Daily Rate (ADR) to be examined at the property level. As illustrated in Figure 3, the representative hotel generated a predicted ADR of 95.591 USD, with the final prediction resulting from the combined influence of both positive and negative feature contributions.

The local explanation indicates that Local Competitive Position provided the largest negative contribution to the predicted ADR, followed by Star Rating, Hotel Type, Room Total, and Site Character: Nature. Conversely, Room

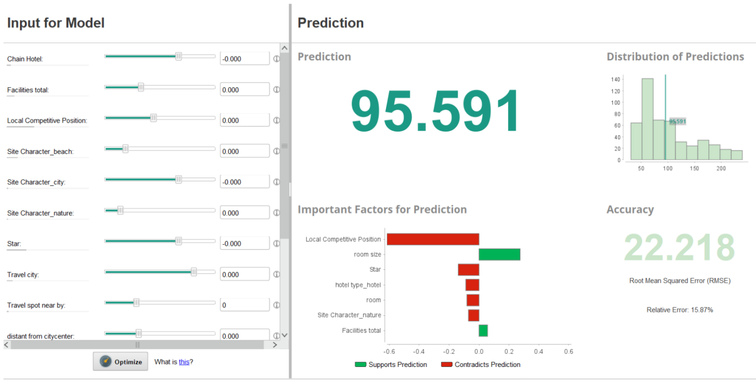


Figure 3. Baseline model simulation and local feature importance analysis.

Size exerted the strongest positive contribution, with additional positive effects from Facilities Total. These results demonstrate that the predicted ADR was determined by the combined influence of contextual market information together with property-specific characteristics rather than by any individual attribute in isolation. Consistent with the global feature importance analysis presented in Section 4.4, Local Competitive Position remained the most influential contextual variable in this representative prediction, highlighting its importance at both the global model level and the individual prediction level. The local surrogate model achieved a Root Mean Square Error (RMSE) of 22.218 USD and a Relative Error of 15.87%, indicating that the surrogate model closely approximated the behavior of the Random Forest model for this individual observation and therefore provided a reliable local explanation.

To further investigate the local response of the Random Forest model, a feature perturbation analysis was conducted within the Model Simulator framework. The simulation examined how modifying the value of the highly influential Local Competitive Position variable affected both the predicted Average Daily Rate (ADR) and the corresponding local feature contributions while all remaining variables were held constant. This analysis provides additional insight into the model's localized decision behavior under a controlled feature perturbation.

As illustrated in Figure 4, perturbing Local Competitive Position resulted in a predicted ADR of 57.535 USD. Relative to the baseline prediction presented in Figure 3, the contribution profile of several variables changed substantially. Room Size became the dominant positive contributor, followed by Room Total, Facilities Total, Hotel Type: Resort, Chain Hotel, and Travel Spot Nearby, whereas Star Rating remained the principal negative contributor. These changes indicate that modifying the contextual market benchmark represented by Local Competitive Position alters the relative contribution of hotel-specific characteristics within the local prediction.

The local surrogate model retained a Root Mean Square Error (RMSE) of 22.218 USD and a Relative Error of 15.87%, indicating that the surrogate explanation continued to closely approximate the behavior of the Random Forest model under the perturbed scenario. Collectively, the perturbation analysis demonstrates that the prediction is determined by the interaction between the contextual market benchmark and property-specific characteristics,While providing an interpretable explanation of how individual feature contributions are redistributed following a controlled change in a key predictor.

To further investigate the combined influence of contextual market conditions and hotel-specific characteristics, a joint multi-attribute perturbation analysis was conducted using the Model Simulator framework. In this analysis, Local Competitive Position was simultaneously changed to $-1$, while Room Size, one of the most influential physical attributes identified in the global feature importance analysis, was increased by one unit. This controlled perturbation was designed to examine how simultaneous changes in both contextual and physical attributes influence the local prediction and feature contribution profile. As illustrated in Figure 5, the combined perturbation increased the predicted Average Daily Rate (ADR) to 127.232 USD. Compared with the previous simulation scenarios, the local explanation exhibited a substantially different contribution profile. Room Size became the dominant positive contributor, followed by Facilities Total and Site Character: Beach, indicating that physical hotel characteristics assumed greater importance under the simulated scenario. Conversely, Room Total, Star Rating,

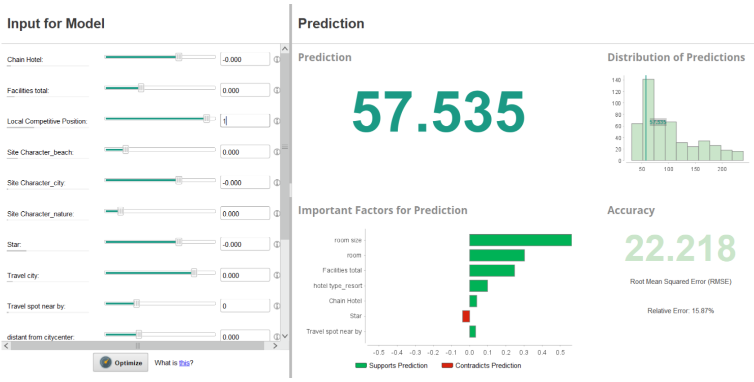


Figure 4. Multi-attribute topology shift under continuous perturbation of Local Competitive Position (+1)

Hotel Type: Hotel, and Site Character: Nature contributed negatively to the prediction, reflecting a redistribution of feature contributions following the simultaneous modification of multiple influential variables. The local surrogate model again maintained a Root Mean Square Error (RMSE) of 22.218 USD and a Relative Error of 15.87%, indicating that the surrogate explanation continued to provide a close approximation of the Random Forest prediction under the joint perturbation scenario. Together with the previous simulations, these results demonstrate that the Random Forest model captures complex interactions between contextual market conditions and hotel-specific characteristics, while the Model Simulator provides an interpretable explanation of how these interactions influence individual predictions.

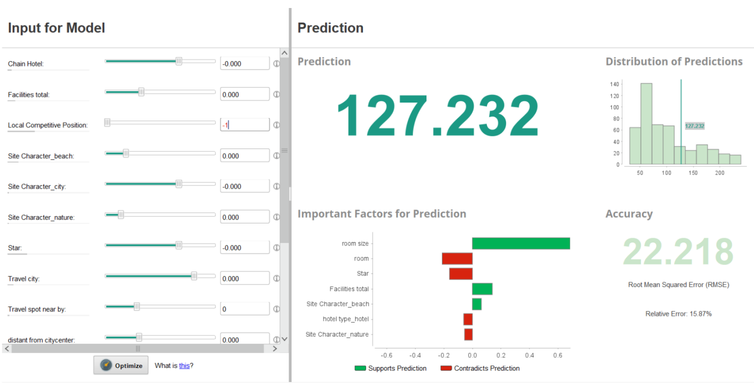


Figure 5. Multi-attribute topology reconfiguration and upward prediction surge under joint perturbation with minimized Local Competitive Position (-1) and elevated room size (+1).

Collectively, the local explanation and perturbation analyses demonstrate an important characteristic of the Random Forest model. Rather than relying on a fixed linear relationship between individual predictors and hotel pricing, the model captures complex interactions among contextual market conditions and property-specific characteristics. As different simulation scenarios were introduced, the relative contribution of individual variables changed accordingly, indicating that the predicted ADR was determined by the combined influence of multiple interacting attributes rather than by any single feature in isolation.

Across all simulation scenarios, Room Size, Facilities Total, Site Character: Beach, Star Rating, and Local Competitive Position consistently emerged as the principal contributors to the local predictions, although their relative influence varied according to the simulated feature values. The local surrogate model consistently maintained a Root Mean Square Error (RMSE) of 22.218 USD and a Relative Error of 15.87%, indicating that the surrogate explanation closely approximated the behavior of the Random Forest model throughout the analyses. These findings demonstrate that the proposed explainability framework provides consistent and interpretable local explanations for understanding how contextual market factors and hotel-specific characteristics jointly influence individual ADR predictions.

### *4.6. Expert Validation of the Machine Learning Findings*

To validate the practical applicability and external reliability of the empirical findings, the quantitative and qualitative outputs derived from the machine learning models were subjected to a formal evaluation by an expert panel comprising eight industry executives (n = 8), whose comprehensive professional profiles are detailed in Table 3. The panel consists of high-level hospitality decision-makers, including hotel CEOs, General Managers of international 5-star chains, and division directors holding between 12 and 25 years of specialized operational experience.

Table 8 presents the quantitative evaluation results generated from the 5-point Likert scale expert assessment. Both primary evaluation criteria comfortably exceeded the preestablished validity threshold of $\geq 3.500$, signifying a strong consensus regarding the model's performance. Specifically, the Reliability of research findings achieved a high mean score of 3.830, while the Practical applicability in the hotel business context scored a mean of 3.910. These parameters indicate that the predictive architecture is regarded as highly trustworthy and practically deployable by industry professionals.

Table 8. Expert Validation Results and Recommendations

| Evaluation Criteria | Mean Score | Interpretation | Expert Recommendations |
|---|---|---|---|
| 1. Reliability of research findings | 3.830 | High | The research findings are applicable for preliminary feasibility studies in hotel development. However, it should be noted that there are numerous external factors beyond the scope of this study that also influence actual business decisions. |
| 2. Practical applicability in hotel business context | 3.910 | High | Findings are applicable for preliminary project feasibility assessment, with acknowledgment that additional external factors influence actual business decisions. |

In addition to the quantitative survey results, structured interviews were conducted to examine whether the feature importance rankings generated by the Random Forest model were consistent with professional experience in the Thai hotel industry. Overall, the expert panel indicated that the relative importance ranking is broadly consistent with their professional experience in the Thai hotel industry. The experts noted that the most influential contextual market variable in the model corresponds to a hotel's competitive pricing position within its local market rather than its absolute room price. They further explained that hotels operating in different provincial markets are subject to varying economic conditions, tourism demand, and competitive environments. As a result, comparing a hotel's room price against comparable four- and five-star properties within the same province provides relevant contextual information for pricing considerations.

The experts also agreed that Star Rating (Weight = 0.796) is identified as one of the most influential predictors of hotel ADR in the model because it reflects the overall service standard and market positioning of a property. Likewise, Room Size (Weight = 0.741) was regarded as an important physical attribute influencing guests' value perception and pricing potential. Site Character: Beach (Weight = 0.552) was consistently identified as an important location-related attribute, particularly for destination markets where beachfront properties typically command higher room prices.

Regarding hotel facilities, the panel indicated that the provision of amenities represented by Facilities Total (Weight = 0.420) is associated with higher perceived customer value and property competitiveness. However, several experts also noted that qualitative aspects not captured in the current dataset, such as room design, maintenance quality, cleanliness, and service delivery, may further influence pricing decisions in practice. These observations suggest that incorporating additional quality-related indicators may further improve future predictive models.

## 5. Discussion

The empirical feature importance topology calculated by the machine learning algorithm offers a data-driven perspective that bridges established hospitality frameworks with localized structural insights within the Thai hotel market.

### 5.1. Local Competitive Position and Hotel Reputation

The feature importance analysis identified Local Competitive Position as the most influential predictor of hotel ADR, indicating that a hotel's pricing position relative to comparable four- and five-star properties within the same provincial market provides important contextual information for price prediction. Unlike absolute room price, this variable reflects a hotel's relative competitive position within its local market, allowing the model to account for regional differences in pricing environments across Thailand. This finding is consistent with previous studies emphasizing the importance of competitive positioning in hotel pricing decisions. For example, [75] demonstrated that hotels maintaining room rates above those of their direct competitors achieved superior revenue performance, while [74] highlighted that understanding the local competitive landscape is essential for strategic hotel positioning. Collectively, these findings suggest that hotel pricing is influenced not only by property characteristics but also by a hotel's relative position within its surrounding competitive market. Among the intrinsic hotel attributes, Star Rating remained the second most influential variable. This finding supports previous hospitality research indicating that official quality classifications continue to serve as important indicators of service standards and pricing capability [1]. Similarly, Chain Hotel demonstrated a positive contribution, suggesting that internationally recognized brands may benefit from stronger consumer confidence and standardized service quality, thereby supporting higher room prices [56, 5].

### 5.2. Location Attributes and Environmental Premium

The feature importance analysis indicates that Site Character: Beach (Weight = 0.552) was considerably more influential than Travel Spot Nearby (Weight = 0.137) and other location-related variables in predicting hotel ADR. This finding suggests that the intrinsic environmental characteristics of a hotel's location may exert a stronger influence on room pricing than simple proximity to tourist attractions. While previous studies have highlighted accessibility and proximity to major attractions as important determinants of hotel value ([21]), the present results indicate that, within the Thai hospitality context, beachfront locations provide an additional environmental premium that extends beyond locational convenience alone. One possible explanation is that beachfront properties represent a relatively scarce tourism resource and offer unique recreational and landscape characteristics that cannot be replicated by hotels located in urban or general locations. Consequently, guests may perceive beachfront hotels as providing greater experiential value, allowing these properties to command higher room rates. This interpretation is further supported by the expert panel, which consistently identified beachfront location as one of the strongest determinants of pricing in Thailand's destination-oriented tourism markets.

### 5.3. Physical Attributes and Pricing Performance

The feature importance analysis identified Room Size (Weight = 0.741) as one of the most influential physical attributes affecting hotel ADR, following Local Competitive Position and Star Rating. This finding suggests that room size remains an important structural characteristic in explaining price variation among four- and five-star hotels in Thailand. Larger rooms generally provide greater functional space and are commonly associated with higher-priced accommodation categories. The expert panel further noted that, in practice, room size should not be interpreted solely in terms of floor area. Instead, larger rooms are often accompanied by higher-quality layouts, improved guest comfort, and enhanced overall room quality in upscale hotels. These observations represent practical industry perspectives rather than direct outputs of the machine learning model and help explain why Room Size emerged as a highly influential predictor. The model interpretability analysis further demonstrated that Room Size consistently acted as an important local contributor under different simulation scenarios. When Local Competitive Position was perturbed, Room Size remained one of the principal variables supporting the model prediction, indicating that physical hotel characteristics continue to contribute substantially to pricing decisions even after changes in market positioning. Similarly, Facilities Total provided additional explanatory value, suggesting that physical amenities complement the structural characteristics of the property in determining ADR.

Overall, these findings indicate that hotel pricing is influenced by both market positioning and physical property characteristics. While Local Competitive Position captures the external market context, Room Size and Facilities Total represent important intrinsic attributes that contribute to guests' perceived value and support higher room prices.

### *5.4. Machine Learning Model Performance*

The comparative evaluation of the five predictive algorithms demonstrates that model performance depends not only on algorithmic complexity but also on the characteristics of the available data. Among the evaluated techniques, Random Forest achieved the highest predictive accuracy ($R^2$ = 0.606; RMSE = 40.99 USD), followed by Multiple Linear Regression, Gradient Boosting, Decision Tree, and Deep Learning. These findings indicate that tree-based ensemble learning provides the most effective framework for predicting hotel Average Daily Rate (ADR) using structured cross-sectional hotel attributes.

The strong performance of Random Forest can be attributed to its ability to capture nonlinear relationships and complex interactions among hotel characteristics while remaining robust to multicollinearity, outliers, and heterogeneous variable distributions. Unlike individual Decision Trees, the ensemble bagging mechanism reduces prediction variance by aggregating multiple decision trees, thereby improving model stability and generalization. This finding is consistent with previous hospitality research, particularly [65], who likewise reported superior predictive performance of Random Forest for hotel price prediction using OTA-derived hotel characteristics.

Interestingly, Multiple Linear Regression achieved performance comparable to the more advanced machine learning models ($R^2$ = 0.575), suggesting that a substantial proportion of ADR variation can still be explained by approximately linear relationships among hotel characteristics. Nevertheless, the superior performance of Random Forest indicates that nonlinear interactions among variables such as Local Competitive Position, Star Rating, Room Size, and Site Character contribute additional explanatory power beyond that captured by a conventional linear model.

From a methodological perspective, the differences in predictive performance across models can be further interpreted through the bias–variance tradeoff and each model's capacity to capture interaction effects. Linear Regression exhibits relatively high bias due to its assumption of linearity and additive effects, limiting its ability to represent complex nonlinear interactions among hotel attributes. In contrast, tree-based ensemble methods such as Random Forest and Gradient Boosting reduce bias by recursively partitioning the feature space and implicitly capturing higher-order interactions without requiring explicit functional specification. This structural flexibility allows these models to better represent the heterogeneous and context-dependent nature of hotel pricing behavior. Deep Neural Networks, while theoretically capable of representing complex nonlinear functions, may exhibit higher variance in small to medium-sized tabular datasets, which can lead to suboptimal generalization performance compared to ensemble tree-based approaches.

By contrast, Deep Learning produced the lowest predictive performance among the evaluated techniques ($R^2$ = 0.473). This outcome is theoretically consistent with previous machine learning studies demonstrating that deep neural networks generally require substantially larger datasets to learn stable hierarchical feature representations effectively. Given the relatively modest sample size of 500 hotels and the structured tabular nature of the predictor variables, the present dataset is more suitable for tree-based ensemble methods than deep neural networks. Under these conditions, Random Forest is able to achieve better generalization while avoiding the overfitting challenges frequently encountered by deep learning models trained on limited tabular data.

Overall, these findings suggest that algorithm selection should be guided by the characteristics of the available dataset rather than model complexity alone. For cross-sectional hotel pricing studies based on structured property-level attributes, tree-based ensemble methods such as Random Forest appear to provide the most reliable balance between predictive accuracy, model stability, and interpretability. Future research employing substantially larger datasets or incorporating dynamic temporal variables, including booking lead time, occupancy, and seasonal demand, may provide a more appropriate environment for advanced deep learning architectures to demonstrate their full predictive potential. From an applied perspective, the proposed machine learning framework can be used as a decision-support tool for hotel pricing strategy and market benchmarking. In particular, the model enables estimation of expected Average Daily Rate (ADR) based on observable hotel attributes and local market conditions, while the feature importance results provide additional interpretive value for understanding relative pricing drivers. The dominant role of Local Competitive Position (LCP) further suggests that hotel pricing performance is more effectively interpreted through a relative market benchmarking perspective rather than isolated property-level characteristics alone.

## 6. Conclusions and Recommendations

### *6.1. Conclusion*

This study demonstrates the applicability of machine learning techniques for predicting hotel Average Daily Rate (ADR) using static property-level attributes collected from an Online Travel Agency (OTA) platform. Five predictive models were evaluated, namely Multiple Linear Regression, Decision Tree, Gradient Boosting, Deep Learning, and Random Forest. Among these models, Random Forest achieved the best predictive performance, with an $R^2$ of 0.606 and an RMSE of 40.99 USD, indicating its ability to capture both linear and nonlinear relationships among hotel characteristics.

A total of fourteen variables representing Contextual Market Factors, Hotel Reputation, Location Attributes, and Physical Attributes were identified as significant predictors of hotel ADR. Feature importance analysis revealed that Local Competitive Position, Star Rating, and Room Size were the three most influential variables, followed by Site Character: Beach and Facilities Total. These findings suggest that hotel pricing is jointly influenced by the competitive conditions of the local market as well as intrinsic characteristics of individual properties. In particular, the importance of Local Competitive Position highlights the value of benchmarking hotels against comparable properties within the same provincial market, enabling the model to account for regional differences in pricing environments across Thailand.

The local interpretability analyses further improved the transparency of the Random Forest model by illustrating how the contribution of individual variables changed under different simulation scenarios. In addition, the expert panel generally agreed that the feature importance rankings were consistent with current hotel pricing practices in Thailand. The experts also emphasized that room size often reflects broader aspects of accommodation quality, while the provision of hotel facilities is associated with the overall value perception of a property rather than determining room prices independently.

Overall, the findings demonstrate that integrating machine learning with interpretable model analysis and expert validation provides a practical framework for understanding hotel pricing behavior. The proposed approach offers useful insights for hotel managers, investors, and researchers seeking to evaluate pricing strategies and identify the relative importance of contextual market conditions and hotel-specific attributes in determining ADR.

### *6.2. Recommendations*

Based on the empirical findings of this study, the following recommendations are offered across five stakeholder groups.

Developers and Entrepreneurs: The proposed model can support feasibility studies for new hotel developments and renovation projects by providing an evidence-based assessment of the relative importance of hotel attributes associated with ADR. Given the strong influence of Local Competitive Position, developers should evaluate planned pricing strategies relative to comparable hotels within the same provincial market rather than relying solely on absolute room price targets. The findings also suggest that investments in hotel quality, including appropriate star classification, room design, and facility provision, should be considered collectively to strengthen the overall market competitiveness of new developments.

Architects and Designers: The feature importance results provide practical guidance for architectural programming and hotel design. In addition to determining appropriate room sizes and facility allocation, design strategies should emphasize room size and spatial configuration, and overall room quality. The expert panel further indicated that room size often represents broader aspects of accommodation quality rather than physical area alone, suggesting that design quality may play an important role in supporting hotel pricing objectives.

Hotel Operators and Revenue Managers: The findings suggest that pricing strategies should be developed with consideration for both hotel-specific characteristics and local market conditions. Because Local Competitive Position was identified as the most influential predictor, hotel operators may benefit from monitoring pricing relative to comparable properties within the same provincial market. Furthermore, maintaining service quality consistent with higher star classifications and enhancing the provision of hotel facilities may strengthen a property's overall value perception and long-term pricing competitiveness.

Researchers: Future research endeavors can expand upon this computational framework in three distinct directions. First, regarding scalability, broadening the web-scraping matrix to encompass multiple synchronized OTA engines or shifting the geographical boundary beyond Thailand will test the cross-border stability of these feature weights. Second, regarding methodological diversity, since standalone Deep Learning underperformed relative to the tree-based ensemble in this cross-sectional context, researchers are encouraged to explore hybrid deep learning architectures (such as TabNet combined with bagging techniques) or integrate neural networks with dynamic temporal inputs, including seasonal demand cycles, real-time occupancy rates, and booking lead times, to push predictive accuracy past static limits. Third, capturing and quantifying the explicit qualitative indicators highlighted by the experts (e.g., quantitative decibel soundproofing metrics, indoor air quality indices) represents a critical path toward constructing next-generation, high-granularity ADR forecasting models.

Government and Tourism Agencies: The findings may assist government agencies in understanding the heterogeneous pricing characteristics of hotels across different regions of Thailand. Such information can support tourism planning by identifying location characteristics associated with higher-value accommodation markets and may contribute to evidence-based understanding regional differences in hotel ADR patterns.

## 7. Limitations

Several limitations should be considered when interpreting the results of this study.

Data source limitation: The dataset was collected exclusively from the Agoda OTA platform, which may not fully capture price variations across different distribution channels. Differences in pricing components such as taxes, discounts, and promotional rates across platforms may also affect the comparability of ADR values. In addition, this study does not distinguish between weekday and weekend rates, nor between promotional and non-promotional rates, due to limitations in the OTA data structure. No explicit outlier removal was applied to the ADR variable. Extreme price values were retained as they reflect observed variability in OTA-based hotel pricing, including seasonal demand variations and heterogeneous market positioning. While the use of 12-month averaging helps reduce short-term price volatility, it does not eliminate underlying seasonal or structural demand effects in the data. Consequently, some degree of price dispersion remains inherent in the dataset. However, ensemble machine learning models help mitigate potential sensitivity to such variations due to their robustness to non-linear and heterogeneous data distributions.

Temporal limitation: Although ADR was computed as a 12 months average to reduce short-term fluctuations, the dataset remains cross-sectional in nature. As a result, the model does not capture dynamic pricing behavior over time or the effects of external shocks. Future research using panel data would allow for a more detailed analysis of temporal pricing dynamics.

Feature limitation: The feature set is limited to structured attributes available from OTA listings. Important qualitative determinants of hotel pricing, such as service quality, aesthetic design, and experiential attributes, are not directly observable in the dataset. In addition, the Facilities Total variable was operationalized as the cumulative number of amenities provided by each property. Although this standardized representation captures the overall breadth of hotel facilities, it does not distinguish the relative importance or market value of individual facility types (e.g., spa, swimming pool, or fitness center). Future studies may investigate weighted facility indices or model individual facility categories separately to better capture their heterogeneous contributions to hotel pricing. Incorporating alternative data sources, such as image-based features or textual reviews, may improve model comprehensiveness.

Model interpretability limitation: While LIME provides useful local explanations for individual predictions, it does not capture global model behavior. In addition, the Random Forest model offers limited global interpretability due to its ensemble structure; therefore, feature importance should be interpreted as relative contributions rather than causal effects.

Generalizability and sample size limitation: The study is based on 500 four- and five-star hotels distributed across Thailand. While the dataset provides nationwide coverage and is sufficient for the applied machine learning models, the findings may not fully generalize to larger datasets or international markets with different economic and regulatory conditions. Future studies are encouraged to validate the proposed framework using larger samples and additional hotel markets to further examine its robustness across different geographical contexts.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found at the Agoda website: https://www.agoda.com. The processed statistical data generated during the study are available within the article or from the corresponding author upon request.

Conflicts of Interest: The authors declare no conflict of interest.

Acknowledgement: This work was supported by the Thammasat University Research Unit for research in Project Development and Innovation in Real Estate Business, Thammasat University.

## REFERENCES


1. B. Agusaj, V. Bazdan, and D. Lujak, "The Relationship between Online Rating, Hotel Star Category and Room Pricing Power," Ekon. Misao Praksa DBK, vol. 26, pp. 189–204, 2017.
2. S. Aksoy and Y. Ozbuk, "Multiple criteria decision making in hotel location: Does it relate to post purchase consumer evaluations?" Tourism Management Perspectives, vol. 22, pp. 73–81, 2017.
3. A. Mattila, "Consumer's Value Judgments: How Business Travelers Assess Luxury-hotel Services," Cornell Hotel and Restaurant Administration Quarterly, 1999. doi: 10.1177/001088049904000121
4. A. Wijoyo, B.A. Septian, S. Intiar, U. A.J and B. A.P, "Qualitative Insights Into the Impact of Room Pricing and Facilities on Hotel Guest Satisfaction," International Journal of Economics, Business Management and Accounting (IJEBMA), vol. 6, no. 2, pp. 131–142, 2024. doi: 10.59890/ijebma.v6i2.2278
5. C. Chen and R. Rothschild, "An application of hedonic pricing analysis to the case of hotel room in Taipei," Tour. Econ., vol. 16, pp. 1–10, 2010.
6. V. Velayutham, G. Chhabra, S. Kumar, A. Kumar, S. Raha, and G. C. Saha, "Analysis of Deep Learning in Real-World Applications: Challenges and Progress," J. Propuls. Technol., vol. 44, 2023. Available: https://doi.org/10.52783/tjjpt.v44.i2.150
7. J. Chica-Olmo, "Effect of monumental heritage sites on hotel room pricing," International Journal of Hospitality Management, vol. 90, 2020. doi: 10.1016/j.ijhm.2020.102640
8. J. D. Egan and K. Nield, "Towards a Theory of Intraurban Hotel Location," Urban Studies, vol. 37, pp. 611–621, 2000.
9. D. Elphick, "Hotel distribution channels: Strategy and examples," SiteMinder, 2025. Available: https://www.siteminder.com/r/hotel-distribution-concepts-best-practice
10. M. Fuchs and K. Weiermair, "Destination benchmarking: An indicator-systems potential for exploring guest satisfaction," Journal of Travel Research, vol. 42, no. 3, pp. 212–225, 2004.
11. F. Khozaei, G. Nazem, T. Ramayah, and S. Naidu, "Factors Predicting Travelers' Satisfaction of Three to Five Star Hotels in Asia, an Online Review," International Journal of Research in Tourism and Hospitality (IJRTH), vol. 2, no. 2, pp. 30–41, 2016.
12. H. Fathima, S. Juveria, S. H. Naaz, I. V. Shashikala, and K. D. Babu, "Future Proofing Real Estate: Machine Learning for Price Predictions," Adv. Comput. Technol. Appl., vol. 9, 2025. Available: https://doi.org/10.5281/zenodo.16596398
13. P. Godinho, P. Phillips, and L. Moutinho, "Hotel Location when competitors may react: A game-theoretic gravitational model," Tourism Management, vol. 69, pp. 384–396, 2018.
14. M. Godolia and A. Sinani, "Assessing Hotel Reservation Forecasting Models to Inform Dynamic Pricing Strategies," in Proc. 10th International Scientific-Business Conference LIMEN, 2024. Available: https://doi.org/10.31410/LIMEN.2024.309
15. T. Hovsepyan, "Deep Learning vs Neural Networks—What's the Difference?" Plat AI, 2025. Available: https://plat.ai/blog/deep-learning-vs-neural-networks-whats-the-difference/
16. M. Huang, W. Zhao, J. Gu, and Y. Lei, "Damage Identification of a Steel Frame Based on Integration of Time Series and Neural Network under Varying Temperatures," Adv. Civ. Eng., vol. 2020, 2020. doi: 10.1155/2020/4284381
17. M. Huang, J. Zhang, J. Hu, and Z. Ye, "Nonlinear modeling of temperature-induced bearing displacement of long-span single-pier rigid frame bridge based on DCNN-LSTM," Case Stud. Therm. Eng., vol. 53, p. 103897, 2024. Available: https://doi.org/10.1016/j.csite.2023.103897
18. D. P. Hutabarat, "Smart system for controlling and monitoring electronic facilities in budget hotel rooms," Indonesian Journal of Electrical Engineering and Computer Science, vol. 30, no. 2, pp. 770–777, 2023. doi: 10.11591/ijeecs.v30.i2.pp770-777
19. I. G. T. Y. P. Jaya, "How room rates drive occupancy: A case study from The Laguna, a Luxury Collection Resort & Spa, Nusa Dua," Journal of Tourism Culinary and Entrepreneurship (JTCE), vol. 3, no. 2, pp. 179–196, 2023. doi: 10.37715/jtce.v3i2.4089
20. A. Israeli, "Star rating and corporate affiliation: their influence on room price and performance of hotel in Israel," Hosp. Manag., vol. 21, pp. 405–424, 2002.

21. G. W. Kim, M. Cho, and A. R. Brymer, "Determinants affecting comprehensive property-level hotel performance: The moderating role of hotel type," Int. J. Hosp. Manag., vol. 31, pp. 404–412, 2013.
22. K. László and H. Ghous, "Efficiency comparison of Python and RapidMiner," Multidiszcip. Tudományok, vol. 10, pp. 212–220, 2020.
23. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, pp. 436–444, 2015.
24. W. Likitanupak, "Hotel's Average Daily Rate (ADR) Prediction Model of 3-5 Star City Hotels and Resort Hotels in Thailand for Customer Booking via Online Travel Agency," Doctoral dissertation, Thammasat University, 2015.
25. P. Lumkam, "Business/Industry Trends 2024-2026: Hotel Business," Krungsri Bank, 2024. Available: P. Lumkam, "Business/Industry Trends 2026-2028: Hotel Business," Krungsri Bank, 2024. Available: https://www.krungsri.com/th/research/industry/industry-outlook/services/hotels/io/hotel-2024-2026
26. V. P. Magnini, J. C. Crotts, and A. Zehrer, "Understanding customer delight: An application of travelblog analysis," Journal of Travel Research, vol. 50, no. 5, pp. 535–545, 2011.
27. L. Masiero, Y. Yang, and T. R. Qiu, "Understanding hotel location preference of customers: Comparing random utility and random regret decision rules," Tourism Management, vol. 73, pp. 83–93, 2019.
28. P. J. Mellinas, L. J. Nicolau, and S. Park, "Inconsistent behavior in online consumer reviews: The effects of hotel attribute rating on location," Tourism Management, vol. 71, pp. 421–427, 2019.
29. P. Morano, F. Tajani, and C. M. Torre, "Artificial intelligence in property valuations. An application of artificial neural networks to housing appraisal," Adv. Environ. Sci. Energy Plan., vol. 39, pp. 23–29, 2015.
30. Y. Narangajavana and B. Hu, "The relationship between the hotel rating system, service quality improvement, and hotel performance changes: a canonical analysis of hotels in Thailand," J. Qual. Assur. Hosp. Tour., vol. 9, pp. 34–56, 2008.
31. N. Nguyen and A. Cripps, "Predicting housing value: A comparison of multiple regression analysis and artificial neural networks," J. Real Estate Res., vol. 22, pp. 313–336, 2001. Available: https://doi.org/10.1080/10835547.2001.12091068
32. M. Oppermann, H. K. Din, and Z. S. Amri, "Urban Hotel Location and Evolution in a Developing Country: The case of Kuala Lumpur Malaysia," Tourism Recreation Research, vol. 21, no. 1, pp. 55–63, 1996. doi: 10.1080/02508281.1996.11014763
33. W. J. O'Neill, "ADR Rule of Thumb," Cornell Hotel Restaur. Adm. Q., 2003. doi: 10.1016/S0010-8804(03)90254-0
34. A. N. Ofori-Boadu, "Exploring Regression Models for Forecasting Early Cost Estimates for High-Rise Buildings," J. Technol. Manag. Appl. Eng., vol. 31, pp. 1–24, 2015.
35. One Nation Travel Experts, "Best Place to Visit in Thailand for First-Timers: Ultimate Guide," 2025. Available: https://www.onenationtravel.com/places-to-visit-in-thailand-for-first-timers
36. R. H. Penner, L. Adams, and W. Rutes, Hotel design: planning and development, Routledge, 2013.
37. R. Pine and P. Qi, "Barriers to hotel chain development in China," International Journal of Contemporary Hospitality Management, 2004. ISSN: 0959-6119.
38. L. Z. Qing and J. C. Liu, "Assessment of the hotel rating system in China," Tour. Manag., vol. 14, pp. 440–452, 1993.
39. J. Rahman, "Simply understanding Artificial Neural Networks (ANN): Deep Learning," Medium, 2024. Available:
40. R. Rajaguru and N. Hassanli, "The Role of Trip Purpose and Hotel star rating on guests' satisfaction and WOM," Int. J. Contemp. Hosp. Manag., vol. 30, 2018. doi: 10.1108/IJCHM-01-2017-0044
41. A. E. Sannindra, E. Tarihan, and R. N. Nugraha, "Analysis of Service Quality to Customer Satisfaction at Ibis Senen Hotel Jakarta," Journal Manajemen Pelayanan Hotel Akademi Komunitas Manajemen Perhotelan Indonesia, vol. 6, no. 2, 2022.
42. L. R. Sestayo, V. M. Bua, and O. L. Gonzalez, "Connection between hotel location and profitability drivers: an analysis of location-specific effects," Current Issues in Tourism, 2018. doi: 10.1080/13683500.2018.1538203
43. T. Shanka and R. Taylor, "An Investigation into the Perceived Importance of Service and Facility Attributes to Hotel Satisfaction," Journal of Quality Assurance in Hospitality & Tourism, vol. 4, no. 3-4, pp. 119–134, 2014. doi: 10.1300/J162v04n03_08
44. J. H. Shi and Q. Su, "Evaluation of hotel service quality based on customer satisfaction," in Proceedings of the 2007 International Conference on Service Systems and Service Management, J. Chen, Ed. Piscataway, NJ: IEEE, 2007, pp. 113–117.
45. S. Shadma and P. Justin, "Examining guests' experience in luxury hotels: evidence from an emerging market," Journal of Marketing Management, 2022. doi: 10.1080/0267257X.2022.2085768
46. N. Shoval, B. McKercher, and A. Birenboim, "Hotel Location and Tourist Activity in Cities," Ann. Tour. Res., vol. 38, pp. 1594–1612, 2011.
47. P. Silvestre, N. Antonio, and P. Carrasco, "Navigating uncertainty: enhancing hotel cancellation predictions with adaptive machine learning," Inf. Technol. Tour., vol. 28, 2026. Available: https://doi.org/10.1007/s40558-025-00349-9
48. I. Soifer, E. Choi, and E. Lee, "Do Hotel Attributes and Amenities Affect Online User Rating Differently across Hotel Star Ratings?" J. Qual. Assur. Hosp. Tour., 2011.
49. J. Suedekum, "Regional costs-of-living with congestion and amenity differences: An economic geography perspective," Ann. Reg. Sci., vol. 43, pp. 49–69, 2009. doi: 10.1007/s00168-007-0201-z
50. The Nation, "Bangkok named the world's second-best city," 2025. Available: https://www.nationthailand.com/thailand/bangkok/40045341
51. K. Tochaiwat and W. Likitanupark, "An Alternative Hotel Pricing Technology: Hedonic Price Model for Pricing Beach Resort Revenue in Thailand," Built, vol. 14, 2019.
52. K. Tochaiwat and A. Budda, "Construction Contract Price Prediction Model for Government Buildings Using a Deep Learning Technique: A Study from Thailand," Buildings, vol. 16, no. 3, p. 651, 2026. Available: https://doi.org/10.3390/buildings16030651
53. R. R. Torrent, M. Fluvia, R. Ballester, A. Salo, E. Ariza, and M. J. Espinet, "The effects of beach characteristics and location with respect to hotel prices," Tour. Manag., vol. 32, pp. 1150–1158, 2011.
54. G. E. Tukenmez, "What are the key attributes behind guest satisfaction in award-winning all-inclusive hotels: An empirical analysis from Türkiye," J. Tour. Econ. Bus. Stud., vol. 7, 2025.
55. J. Villamediana, I. Küster-Boluda, N. Vila-Lopez, G. Piñero, and V. Naranjo, "Variables related to prices for Belgium, France, Germany, Italy, Netherlands, Spain, United Kingdom and United States," Data set, Zenodo, 2024. Available: https://doi.org/10.5281/zenodo.10533156

56. L. Wu, ”The pricing of a brand name product: Franchising in the motel services industry,” J. Bus. Venturing, vol. 14, pp. 87–102, 1998.
57. Z. Xiang and M. Krawczyk, ”What does hotel location mean for the online consumer? Text Analytics Using Online Reviews,” J. Travel Res., vol. 57, pp. 925–939, 2016.
58. Y. Yang, K. K. Wong, and Wang, ”How do hotels choose their location? Evidence from hotels in Beijing,” International Journal of Hospitality Management, 2012. doi: 10.1016/j.ijhm.2011.09.003
59. Y. Yang, H. Luo, and R. Law, ”Theoretical, empirical, and operational models in hotel location research,” International Journal of Hospitality Management, vol. 36, pp. 209–220, 2014.
60. Y. Yang, Z. Mao, and J. Tang, ”Understanding Guest Satisfaction with Urban Hotel Location,” Empirical Research Articles, 2017. DOI:10.1177/0047287517691153
61. X. J. Zhang and H. M. Li, ”Budget hotel customer satisfaction analysis based on SERVQUAL model,” Strategic Management Engineering: Enterprise, Environment and Crisis, vol. 1, pp. 64–68, 2009.
62. Z. Zhou, ”Research on the Spatial Design of resort Hotel from the perspective of Regional culture,” Highlights Art Des., vol. 2, pp. 29–31, 2023. Available: https://doi.org/10.54097/hiaad.v2i1.5192
63. M. T. Ribeiro, S. Singh, and C. Guestrin, ”“Why Should I Trust You?” Explaining the Predictions of Any Classifier,” in Proc. ACM Press the 22nd SIGKDD International Conference, pp. 1135-1144, 2016. https://doi.org/10.1145/2939672.2939778
64. A. Belarmino, Y. Koh, and H. H. Shin, ”Blurred lines: comparing room rates, star-ratings, and user-generated contents for full-service and select-service hotels,” Serv Bus, vol. 16, no. 3, pp. 721–742, 2022. doi: 10.1007/s11628-021-00477-2
65. T. Hu and H. Song, ”Analysis of influencing factors and distribution simulation of budget hotel room pricing based on big data and machine learning from a spatial perspective,” Sustainability, vol. 15, no. 1, p. 617, 2023. Available: https://doi.org/10.3390/su15010617
66. L. Breiman, ”Random Forests,” Machine Learning, vol. 45, pp. 5–32, 2001. Available: https://doi.org/10.1023/A:1010933404324
67. J. H. Friedman, ”Greedy Function Approximation: A Gradient Boosting Machine,” The Annals of Statistics, vol. 29, no. 5, 2000. doi: 10.1214/aos/1013203451
68. M. Zorman, M. M. Štiglic, P. Kokol, and I. Malčić, ”The Limitations of Decision Trees and Automatic Learning in Real World Medical Decision Making,” Journal of Medical Systems, vol. 21, pp. 403–415, 1997. Available: https://doi.org/10.1023/A:1022876330390
69. D. C. Montgomery, E. A. Peck, and G. G. Vining, Introduction to Linear Regression Analysis, 5th ed. Wiley, 2012.
70. B. De Ville, ”Decision trees,” WIREs Comput Stat, vol. 5, pp. 448–455, 2013. Available: https://doi.org/10.1002/wics.1278
71. Y. Zhang, Y. Zhang, and X. Li et al., ”Hospitality service aesthetic of luxury hotel front desk and its effect on consumers’ willingness-to-pay price premium,” Curr Psychol, vol. 43, pp. 23635–23655, 2024. Available: https://doi.org/10.1007/s12144-024-06085-3
72. K. Kirillova and J. Chan, ”“What is beautiful we book”: hotel visual appeal and expected service quality,” International Journal of Contemporary Hospitality Management, vol. 30, no. 3, pp. 1788–1807, 2018. doi: https://doi.org/10.1108/IJCHM-07-2017-0408
73. 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, 2016, pp. 785–794. doi: 10.1145/2939672.2939785
74. R. Zhou, B. Song., G. Li and R. Law, “A data-driven approach to competitor identification and categorization in the hotel industry,” International Journal of Hospitality Management, vol. 133, 2026, Available: https://doi.org/10.1016/j.ijhm.2025.104503
75. C.A. Enz, L. Canina and J.I. Van Der Rest, “Hotel strategic pricing in Europe: a 10-year exploration of competition,” International Journal of Revenue Management, 2016, vol. 9, no. 2-3, pp. 92-107. Available: https://doi.org/10.1504/IJRM.2016.077016
76. H. Zhang, J. Zhang, S. Lu, S. Cheng and J. Zhang, “Modeling hotel room price with geographically weighted regression,” International Journal of Hospitality Management, 2011, vol. 30, Issue 4, pp. 1036-1043. Available: https://doi.org/10.1016/j.ijhm.2011.03.010
77. A. Ampountolas, “Hybrid machine learning approaches for hotel occupancy forecasting: Evaluating gradient boosting and neural networks,” International Journal of Hospitality Management, 2026, vol. 139. Available: https://doi.org/10.1016/j.ijhm.2026.104745
78. S. Kisilevich, D. Keim, L. Rokach, “A GIS-based decision support system for hotel room rate estimation and temporal price prediction: The hotel brokers’ context” Decision Support Systems, 2013, vol 54, Issue 2, pp. 1119-1133. Available: https://doi.org/10.1016/j.dss.2012.10.038
79. R. Razavi and AA. Israeli, “Determinants of online hotel room prices: comparing supply-side and demand-side decisions,” International Journal of Contemporary Hospitality Management, 2019, vol. 31, no. 5, pp. 2149–2168, Available: https://doi.org/10.1108/IJCHM-09-2018-0707
80. S. Kumar Mitra, “Estimating the duration of different seasons and their impact on hotel room prices,” International Journal of Hospitality Management, 2020, vol 90, Available: https://doi.org/10.1016/j.ijhm.2020.102604
81. Pine, B.J. & Phillips, J. (2004). Title of the reference. Journal/Book, volume(issue), pages.