International Journal of Informatics and Data Science Research
ISSN 2997-3961 (Online)
Vol. 3, No. 2, Feb 2026,
Available online at: https://scientificbulletin.com/index.php/IJIDSR

Article

Operationalizing SHAP for Imbalanced No-Show Prediction in Hospitality

[1]I Gusti Ngurah Agung Krishna Aditya, [2]Syadia Nabilah Binti Mohd Safuan, [3]I Nyoman Gede Arya Astawa.
[1] School of Graduate Studies, Management Science and University, Malaysia
[2]Faculty of Information Science & Engineering, Management Science and University, Malaysia
[3]Department of Information and Technology, Bali State Polytechnic, Indonesia

*Corresponding author's email: krishna.aditya125@gmail.com, syadia_nabilah@msu.edu.my, arya_kmg@pnb.ac.id

Article information:
Manuscript received: 10 Feb 2026; **Accepted:** 12 Mar 2026; **Published:** 09 Apr 2026

Abstract: Hotel revenue management relies on accurately forecasting capacity. When a guest books a room, the property commits a perishable asset. If the guest fails to arrive without formally cancelling, the hotel absorbs a complete financial loss for that given night. Identifying these "no-show" bookings represents a difficult classification problem due to the rarity of the behaviour. In this dataset of 87,396 validated records, only 1.58% of reservations resulted in a realized no-show. Standard machine learning models generally fail when confronted with a 62-to-1 class imbalance. They achieve high raw accuracy by predicting that every guest will arrive, which provides zero operational utility. This paper details an XGBoost classification pipeline explicitly optimized to identify the minority class. By adjusting the class weights (scale_pos_weight = 62.477), the model accepted lower absolute precision to achieve a 0.60 recall rate on the minority class, successfully identifying 60% of all no-shows in the test set. The final model returned a ROC-AUC of 0.7406 and an F2-score of 0.1516, reflecting the heavy weighting toward recall. This research then subsequently applied SHAP (SHapley Additive exPlanations) to interpret the algorithmic decision-making process. Permutation importance tests and SHAP matrices confirmed that the absence of a parking space request was the primary predictor of a no-show. The resulting framework provides front-desk teams with clear, explainable thresholds for implementing targeted deposit policies.

Keywords: Machine Learning, XGBoost, Explainable AI, SHAP, Class Imbalance, Capacity Forecasting.

1. Introduction

No-show remains a primary challenge in the hospitality industry, while a canceled reservation provides the sales team with time to resell the inventory, a no-show leaves the room empty, resulting in immediate revenue loss [1]. Hotels attempt to mitigate this financial risk by intentionally overbooking their properties. If statistical models predict a certain volume of cancellations and no-shows, the property sells inventory beyond its actual physical capacity. However, when these forecasting algorithms underestimate arrival rates, the property faces an overcapacity event [2]. Management is then forced to displace arriving guests to competitor hotels, incurring immediate financial penalties and long-term reputational damage. To prevent displacement events, operations teams require predictive frameworks that identify which specific reservations hold the highest probability of resulting in a no-show. This requires overcoming significant class imbalance. In the 87,396 records analyzed in this study, only 1,376 reservations resulted in a no-show, representing a 1.58% base rate. Standard machine learning algorithms optimize for overall accuracy [3]. When provided with a 1.58% event rate, standard models achieve 98.42%

accuracy by simply predicting the majority class for every instance. While mathematically optimal under default parameters, this approach yields no utility for capacity management. This paper constructs a pipeline specifically designed to address this structural imbalance. This research utilized XGBoost, as gradient boosted trees handle non-linear interactions efficiently and allow for explicit penalization of missed minority classes [4]. The framework is configured to strictly optimize for the minority class rather than overall accuracy.

However, predicting risk in isolation is insufficient for physical operations. If an automated system recommends requiring a deposit from a specific guest or prioritizing a reservation for overbooking displacement, human managers require interpretable reasoning. SHapley Additive exPlanations (SHAP) is integrated to translate the XGBoost probability outputs into transparent operational logic [5]. By quantifying the exact risk assigned to specific booking features, such as lead time and daily rate, revenue teams are provided with a mathematically sound justification for operational interventions.

2. Research Methodology

2.1 Pipeline Architecture

To understand the architecture driving this model, it is necessary to map the raw data ingestion through to the final interpretability layer. Property management systems typically generate inconsistent data structures [6]. Our objective was to standardize this input into a mathematically rigorous environment where the classification logic could be isolated for operational analysis.

The analytical pipeline was structured as a sequential funnel. Initially, the raw dataset underwent rigorous hygiene protocols, specifically deduplication, to produce a verified baseline dataset. Following sanitation, the pipeline isolated the target classification, distinctly separating standard check-outs from realized no-shows to determine the structural base rate. Once the minority target was isolated, the continuous variables underwent extreme-value capping and composite feature derivation.

The resulting standardized matrix was then strictly stratified into three distinct mathematical subsets: a primary training set, an isolated validation set for hyperparameter tuning, and a holdout test set to evaluate final generalization. The training and validation sets were directly ingested by the gradient boosted classification algorithm. Following training completion, the model underwent sequential verification. Initial permutation importance tests confirmed systemic validity, allowing the final model parameters and the holdout test set to be passed into the SHAP extraction layer. This terminal stage deconstructed the algorithmic predictions, outputting the final behavioural interaction plots and risk indicators.

2.2 Dataset Composition

This analysis utilized the Hotel Booking Demand dataset [7], originally comprising 119,390 observations and 32 feature variables collected from two property types (a Resort Hotel and a City Hotel). Prior to any preprocessing interventions, the raw 32-feature space encompassed a spectrum of guest demographics, logistical timelines, and financial metrics. The primary feature groups included:

- Temporal Indicators: lead_time, arrival_date_year, arrival_date_month, arrival_date_week_number, and arrival_date_day_of_month.
- Occupancy Logistics: adults, children, babies, stays_in_weekend_nights, and stays_in_week_nights.
- Financial and Classification Parameters: adr (Average Daily Rate), customer_type, market_segment, distribution_channel, meal, and deposit_type.

- Guest Interaction Metrics: required_car_parking_spaces, total_of_special_requests, booking_changes, previous_cancellations, and days_in_waiting_list.
- System Tracking Variables: hotel, agent, company, reserved_room_type, assigned_room_type, reservation_status, and reservation_status_date.

These unstructured variables formed the baseline input matrix before applying standard cleaning protocols or synthesized composite operational metrics.

2.3 Data Preprocessing

The pipeline began with 119,390 raw bookings sourced from hotel property systems. Hospitality datasets frequently contain duplicate entries resulting from network timeouts or redundant API calls from third-party Online Travel Agencies [8]. This research executed a systematic deduplication process that removed over 30,000 invalid records, leaving 87,396 unique reservations. Training the algorithm on redundant data would artificially inflate the performance metrics, compromising its predictive validity.

The target variable required isolated curation. Standard reporting systems frequently group "Canceled" and "No-Show" bookings into a single failure category [9]. In practice, a cancellation provides lead time for inventory resale, whereas a no-show guarantees revenue loss. Grouping these distinct events obscures the predictive variables. Consequently, we defined the binary target strictly as a comparison between Check-Out (Class 0) and No-Show (Class 1). This isolation produced the 1.58% event rate observed in the dataset.

Missing data was addressed semantically. If an agent or company ID was absent, we imputed a baseline value of -1 to indicate a direct booking channel rather than applying mean or median replacements. Continuous variables were capped at their 99th percentiles to prevent extreme outliers from distorting the decision boundaries. The Average Daily Rate (ADR) was limited to a maximum of 342.29, and booking lead times were constrained to 480 days. Composite features are derived to evaluate financial burden per occupant:

$$\mathrm{ADR}_{\text{per person}} = \frac{\text{Average Daily Rate}}{\text{Adults} + \text{Children} + \text{Babies}}$$

2.4 Leakage Prevention

The feature space contained columns prone to data leakage. Internal tracking variables like reservation_status_date are generated during the check-in process [10]. Including these metrics allows the algorithm to predict outcomes based on data that remains unavailable at the actual time of booking, reservation_status_date, is_canceled, and country prior to initiating the training phase is removed. The data was distributed into three distinct subsets to ensure rigorous evaluation:

- Training Set: 45,069 samples
- Validation Set: 9,658 samples
- Testing Set: 9,658 samples

Strict stratification across all splits then applied. This maintained the 1.58% no-show event rate proportionately across the training, validation, and testing subsets, ensuring the algorithm evaluated representative distributions in every batch.

2.5 Mathematical Weighting

Gradient boosted tree architectures like XGBoost process tabular data effectively but require explicit weighting to address severe class imbalances [11]. If unweighted, the algorithm achieves 98.42% accuracy by universally predicting check-outs, providing no utility.

The primary configuration relied on calculating a strict penalty ratio via the scale_pos_weight hyperparameter. This ratio assigns elevated mathematical cost to missed minority class instances. Value is calculated directly from the training data distribution:

$$\text{scale\_pos\_weight} = \frac{\sum_{i=1}^{N} \mathbb{I}(\, y_i = 0)}{\max(\sum_{i=1}^{N} \mathbb{I}\,(y_i = 1),1)}$$

Applied to the 45,069 samples in the training set, this formula yielded the following configuration:

$$\text{scale\_pos\_weight} = 62.477$$

By assigning a penalty weight of 62.477, we instructed the XGBoost algorithm that failing to identify a single no-show incurs the same mathematical penalty as misclassifying 62 standard check-outs.

The model utilized 200 base estimators and limited the maximum tree depth to 6 splits to control variance. Training occurred against the 9,658-sample validation set, utilizing an early-stopping trigger of 20 rounds to prevent overfitting. Performance was evaluated exclusively through the optimization of the Area Under the Receiver Operating Characteristic Curve (ROC-AUC), as this metric provides a threshold-independent assessment of the model's ability to rank positive (no-show) instances higher than negatives across all decision boundaries [12].

3. Result

3.1 Final Test Set Performance

Evaluating models trained on highly imbalanced datasets requires specialized classification metrics. Raw accuracy is explicitly deprioritized. In a hospitality context, the cost of a false negative (failing to predict a no-show, leading to empty inventory) is significantly higher than the cost of a false positive (incorrectly categorizing an arriving guest as high-risk). This research utilized the F2-Score, which weights recall twice as heavily as precision, to better align the model evaluation with operational priorities.

The XGBoost model achieved a ROC-AUC score of 0.7406 and an F2-Score of 0.1516 on the holdout test set of 9,658 previously unseen records.

Table Error! No text of specified style in document..**1** Classification Report.

Metric	Precision	Recall	F1-Score	Support
Class 0 (Check-Out)	0.99	0.76	0.86	9,506
Class 1 (No-Show)	0.04	0.60	0.07	152
Macro Average	0.51	0.68	0.47	9,658
Weighted Average	0.98	0.76	0.85	9,658

While the absolute precision for Class 1 is statistically low (0.04), this outcome is a direct and intended result of the scale_pos_weight hyperparameter configuration. The objective was to prioritize recall. The algorithm successfully identified 60% of all true no-shows. The operational implication is that management systems will generate a high volume of false positives. However, requiring flagged guests to review automated confirmation policies or post refundable deposits mitigates the financial risk without denying service to arriving customers.

3.2 Permutation Importance Validation

Prior to interpreting the model through SHAP analysis, a permutation importance verification is executed. This process randomized individual data columns to measure the resulting degradation in the ROC-AUC score [13]. Significant decreases in accuracy following the randomization of a seemingly irrelevant variable indicate potential data leakage or spurious correlations.

This research applied a custom scoring function to isolate drops in the ROC-AUC metric rather than measuring divergence in baseline accuracy.

Table Error! No text of specified style in document..**2** Top 10 Features by Permutation Importance.

Rank	Feature	Mean ROC-AUC Decrease	Standard Deviation
1	required_car_parking_spaces	0.0885	0.0200
2	hotel	0.0549	0.0131
3	agent	0.0375	0.0090
4	lead_time	0.0346	0.0406
5	arrival_date_day_of_month	0.0305	0.0130
6	arrival_date_week_number	0.0300	0.0117
7	stays_in_week_nights	0.0283	0.0094
8	customer_type	0.0274	0.0161
9	adr_per_person	0.0270	0.0114
10	adr	0.0241	0.0213

Parking space requirements demonstrated the highest importance, decreasing the ROC-AUC by nearly 9 points when randomized. The primary predictive variables correlated with logical booking patterns and temporal distance rather than system processing artifacts. This structural integrity validated the progression to SHAP dependency modeling.

4. Discussion

4.1 Model Interpretability

Predicting no-show probabilities is mathematically functional when models strictly penalize minority class classification errors. However, understanding the underlying causality is essential for production deployment. Providing operational management with isolated probability thresholds lacks the necessary context to warrant direct customer intervention [14]. If staff must evaluate a booking to enforce a deposit or cancellation policy, they require interpretable rationale.

Machine learning models within hospitality operations generally operate as opaque systems providing limited explanatory output. Global feature importance parameters, such as the permutation rankings established in the results section, quantify systemic relevance but fail to detail whether a specific feature increased or decreased risk for an isolated prediction. This research applied SHapley Additive exPlanations (SHAP) to transform the classification probabilities into transparent operational data.

The SHAP framework isolates the local impact of all input features across individual predictions, calculating exact directional influence. Mathematically, SHAP distributes the final probability prediction among the constituent features based on coalitional game theory. The model output $\hat{y}$ represents the sum of the base expected value ϕ_0 and the individual feature contributions ϕ_i:

$$\hat{y}(x) = \phi_0 + \sum_{i=1}^{M} \phi_i (x)$$

By calculating these parameters across the 9,658 test samples, we generated an explanatory matrix of no-show correlations.

4.2 Impact of Parking Requests

The global summary matrix substantiated the permutation test outputs. A primary correlation between temporal distance, low financial commitment, and the absence of logistical planning was observed.

The most prominent predictive variable was the required_car_parking_spaces parameter. When a guest requested a parking space, the associated SHAP value trended sharply negative, correlating with a lower probability of a no-show. A parking request serves as a quantifiable indicator of logistical preparation.

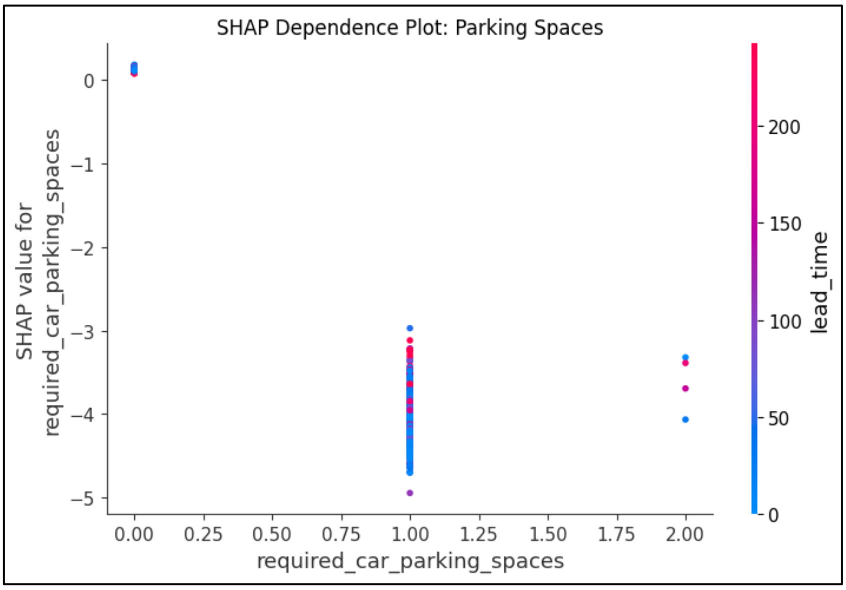

Figure Error! No text of specified style in document.**.1:** SHAP Dependence Plot: Parking Spaces.

Logging a vehicle reservation strongly denotes active physical planning for the upcoming itinerary. The absence of a parking request correlates with a higher baseline risk, indicating that the booking may represent a speculative option rather than a finalized travel transition. While variables such as public transit utilization alter the baseline volume of parking requests at urban properties, the underlying statistical correlation between active logistical requests and arrival probability remains robust.

4.3 Role of Financial Commitment

Initial hypotheses assumed that premium room rates would increase cancellation and abandonment probabilities due to financial burden. The empirical data contradicted this assumption. Bookings associated with exceptionally low daily rates exhibited the highest probability of resulting in a no-show.

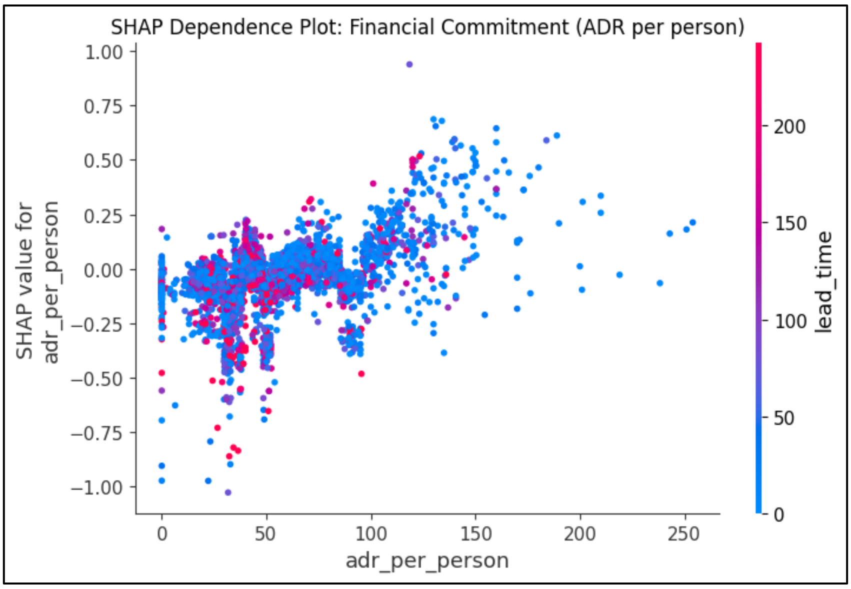

Figure Error! No text of specified style in document..**2:** SHAP Dependence Plot: Financial Commitment (ADR per person)

When the Average Daily Rate (ADR) per person nears zero, the predictive SHAP value increases significantly. This dynamic persists whether the rate represents a strictly budget booking or an artificially lowered rate due to promotional codes and loyalty redemptions.

This behaviour illustrates an absence of financial commitment. Highly discounted inventory requires minimal initial investment from the guest, significantly reducing the financial consequence of abandoning the reservation. Conversely, standard retail or premium reservations correlate with high arrival rates. The financial penalty associated with forfeiting a standard block booking enforces behavioural compliance. If a user commits significant capital, they generally either check in or execute a standard cancellation to initiate a refund. Permitting highly flexible cancellation policies on heavily discounted inventory directly correlates with operational volatility.

4.4 Lead Time and Temporal Factors

The absence of financial commitment interacts directly with the length of the booking window. Reservations created 300 days prior to arrival are statistically less stable due to external scheduling variables and flight disruptions.

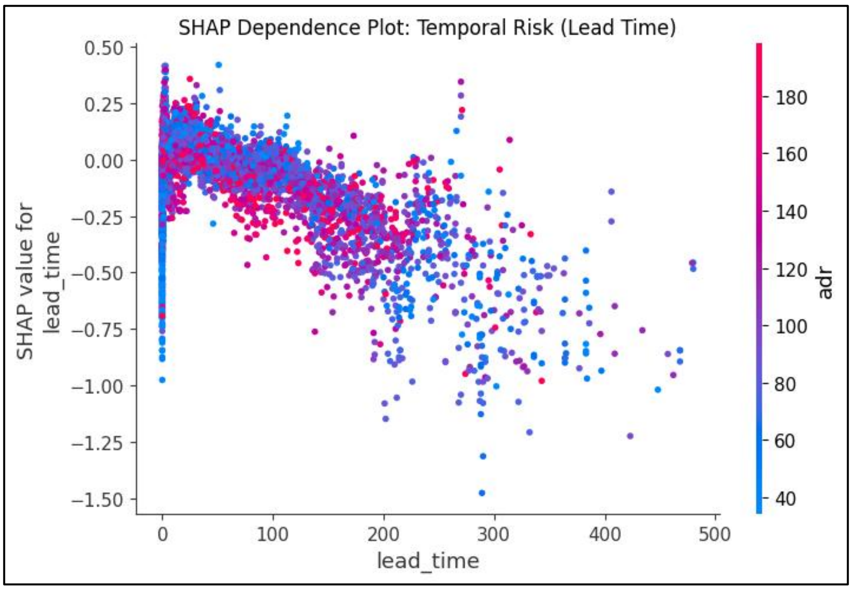

Figure Error! No text of specified style in document..**3:** SHAP Dependence Plot: Temporal Risk (Lead Time).

The SHAP temporal dependency plot highlights a specific boundary logic. Extended lead times display high correlation with no-show behaviour only when paired with low average daily rates. Consumers reserving heavily discounted inventory several months in advance frequently utilize the booking as a rate-lock mechanism while continuously comparing alternative options over the duration of the booking window.

Third-party processing channels, specifically Online Travel Agencies (OTAs), actively facilitate this pattern. Digital travel platforms frequently permit users to secure multiple refundable bookings concurrently across competing properties. Therefore, a reservation made six months prior at a premium price point indicates stability, whereas an extended-lead reservation secured at a minimal price displays a high correlation with future abandonment.

4.5 Stay Duration and Property Type Dynamics

The intersection of booking duration and physical property type provides a secondary, critical predictive dimension. While standard length-of-stay metrics often yield linear baseline correlations, the SHAP dependency evaluation reveals a distinct behavioural bifurcation based on the specific hotel classification.

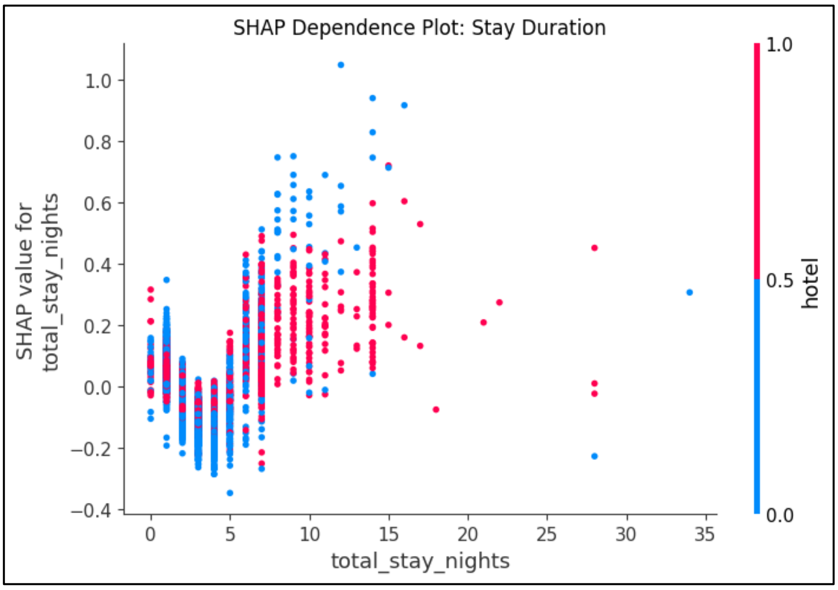

Figure Error! No text of specified style in document..**4:** SHAP Dependence Plot: Stay Duration.

For short-duration bookings (under five nights), the associated SHAP values remain generally negative or neutral, indicating stable arrival probabilities. However, as the duration extends beyond seven nights, the no-show probability increases substantially, segmented heavily by property type. Extended stays at resort properties exhibit standard baseline risk, as two-week vacations represent typical consumer behaviour in that segment. Conversely, extended bookings at city properties generate exponentially higher SHAP risk values.

A fourteen-night reservation at a transient urban property often represents anomalous consumer behaviour. These extended urban bookings frequently serve as speculative placeholders for visa applications or alternative immigration requirements rather than genuine travel itineraries. By capturing this interaction, the model effectively identifies bookings that deviate from the systemic norms of the specific property type, allowing management to apply strict verification requirements to statistically abnormal extended stays.

4.6 Operational Intervention Policies

Applying these algorithmic matrices to physical management systems provides operational teams with structured frameworks for intervention. Rather than initiating customer contact based on heuristic evaluations, management structures can rely on automated policies derived from intersectional risk features. The analysis isolates three primary behavioural vectors: extended temporal lead times, low financial commitment, and the absence of logistical planning interactions. The intersection of these variables dictates when intervention is statistically justified.

If a system detects a reservation scheduled more than 200 days in advance, featuring a deeply discounted rate, and logging no additional logistical requests, the property management software can automatically issue a verification or deposit requirement protocol. This shifts the operational burden from the front desk staff to an automated verification process. Given that the objective is validating arrival intent, enforcing nominal deposit structures effectively filters high-risk speculative bookings. Future analyses should evaluate external parameters including weather forecasting algorithms and aviation delay metrics to measure environmental variance against baseline financial commitments. However, relying on financial and temporal indicators currently provides the most statistically reliable determinant of arrival probability.

5. Conclusion

Algorithmic attempts to forecast hotel capacity failure frequently prioritize raw modeling accuracy over operational utility [15]. This methodology was engineered specifically to adhere to the operational constraints of inventory management. While the final model produced a statistically low precision rate of 4% regarding the positive class, this tradeoff successfully classified 60% of all realized no-show events within a highly imbalanced dataset characterized by a 1.58% base rate. This emphasis on recall over precision is required for systems operating in production environments. Enforcing automated deposit verification protocols for a higher volume of false positives remains computationally and financially more efficient than processing unexpected inventory displacement due to unforecasted overcapacity.

Translating the XGBoost non-linear probability outputs through SHAP values provided interpretable logic for administrative review. The researcher then identified the underlying behaviours associated with no-show realization: extended temporal lead distances, low financial consequence, and the absence of verifiable physical planning parameters. Deploying this architecture requires integrating the calculated SHAP thresholds directly with existing Property Management Systems (PMS). By triggering automated verification requests strictly upon the structural intersection of these specific risk markers, hospitality operations can begin establishing data-driven deposit parameters and reduce their reliance on heuristic overbooking quotas.

6. Declaration

6.1 Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

6.2 Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

6.3 Data Availability Statement

The analysis in this study was conducted using the publicly available Hotel Booking Demand dataset [7].

6.4 Author Contribution

All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by I Gusti Ngurah Agung Krishna Aditya. Supervision, methodology validation, and critical review were conducted by Syadia Nabilah Binti Mohd Safuan and I Nyoman Gede Arya Astawa. The first draft of the manuscript was written by I Gusti Ngurah Agung Krishna Aditya, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

REFERENCES

[1] Q. Zhai, Y. Tian, J. Luo, and J. Zhou, "Hotel overbooking based on no-show probability forecasts," *Comput. Ind. Eng.*, vol. 180, p. 109226, Jun. 2023, doi: 10.1016/J.CIE.2023.109226.

[2] M. Dudziak, "How to Reduce No-Shows and Last-Minute Cancellations | Hotel Tech Insight." Accessed: Jan. 09, 2026. [Online]. Available: https://hoteltechinsight.com/2026/03/17/how-to-reduce-hotel-no-shows-cancellations/

[3] Andy Hermawan, Aji Saputra, Nabila Lailinajma, Reska Julianti, Timothy Hartanto, and Troy Kornelius Daniel, "Predicting Hotel Booking Cancellations Using Machine Learning for Revenue Optimization," *Router : Jurnal Teknik Informatika dan Terapan*, vol. 3, no. 1, pp. 37–48, Mar. 2025, doi: 10.62951/router.v3i1.400.

[4] S. Yang, "Predicting hotel booking cancellations using the XGBoost algorithm," Tilburg University, Tilburg, 2022.

[5] Z. Liu, K. W. De Bock, and L. Zhang, "Explainable profit-driven hotel booking cancellation prediction based on heterogeneous stacking-based ensemble classification," *Eur. J. Oper. Res.*, vol. 321, no. 1, pp. 284–301, Feb. 2025, doi: 10.1016/J.EJOR.2024.08.026.

[6] Y. A. Singgalen, "Big Data in Tourism and Hospitality Industry: Predictive Analytics of Hotel Room Trends," *Indonesian Journal of Tourism and Leisure*, vol. 06, no. 1, p. 2025, 2025, doi: 10.36256/ijtl.v6i1.474.

[7] N. Antonio, A. de Almeida, and L. Nunes, "Hotel booking demand datasets," *Data Brief*, vol. 22, pp. 41–49, Feb. 2019, doi: 10.1016/j.dib.2018.11.126.

[8] N. Antonio, A. De Almeida, and L. Nunes, "An automated machine learning based decision support system to predict hotel booking cancellations," *Data Sci. J.*, vol. 18, no. 1, 2019, doi: 10.5334/dsj-2019-032.

[9] P. Silvestre, N. Antonio, and P. Carrasco, "Navigating uncertainty: enhancing hotel cancellation predictions with adaptive machine learning," *Information Technology & Tourism 2025 28:1*, vol. 28, no. 1, pp. 9-, Dec. 2025, doi: 10.1007/S40558-025-00349-9.

[10] M. Balawejder, "Monitoring a Hotel Booking Cancellation Model: Creating Reference and Analysis Set." Accessed: Jan. 09, 2026. [Online]. Available: https://www.nannyml.com/blog/nannyml-with-sagemaker-part-1

[11] C. Wang, C. Deng, and S. Wang, "Imbalance-XGBoost: leveraging weighted and focal losses for binary label-imbalanced classification with XGBoost," *Pattern Recognit. Lett.*, vol. 136, pp. 190–197, Aug. 2020, doi: 10.1016/J.PATREC.2020.05.035.

[12] E. Richardson, R. Trevizani, J. A. Greenbaum, H. Carter, M. Nielsen, and B. Peters, "The ROC-AUC Accurately Assesses Imbalanced Datasets," 2023, doi: 10.2139/SSRN.4655233.

[13] S. Janitza, C. Strobl, and A. L. Boulesteix, "An AUC-based permutation variable importance measure for random forests," *BMC Bioinformatics*, vol. 14, p. 119, Apr. 2013, doi: 10.1186/1471-2105-14-119.

[14] R. Dai, "The impact of customer perceived value on relationship quality - An empirical study of Home Inns based on SEM, SVM and SHAP value," *Proceedings of the 2nd Guangdong-Hong Kong-Macao Greater Bay Area International Conference on Digital Economy and Artificial Intelligence, DEAI 2025*, pp. 1846–1852, Jul. 2025, doi: 10.1145/3745238.3745526;PAGEGROUP:STRING:PUBLICATION.

[15] D. Contessi, L. Viverit, L. N. Pereira, and C. Y. Heo, "Decoding the future: Proposing an interpretable machine learning model for hotel occupancy forecasting using principal component analysis," *Int. J. Hosp. Manag.*, vol. 121, p. 103802, Aug. 2024, doi: 10.1016/J.IJHM.2024.103802.