IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 9 January 2025, accepted 21 January 2025, date of publication 29 January 2025, date of current version 12 February 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3536094

RESEARCH ARTICLE

Machine Learning in Hospitality: Interpretable Forecasting of Booking Cancellations

ISMAEL GÓMEZ-TALAL[1,2], (Member, IEEE), MANA AZIZSOLTANI[3,4,5], PILAR TALÓN-BALLESTERO[2], AND ASHOK SINGH[3]

[1]Department of Signal Theory and Communications and Telematic Systems and Computation, Rey Juan Carlos University, Fuenlabrada, 28943 Madrid, Spain
[2]Department of Business and Management, Rey Juan Carlos University, Fuenlabrada, 28943 Madrid, Spain
[3]William F. Harrah College of Hospitality, University of Nevada, Las Vegas, Las Vegas, NV 89154, USA
[4]International Gaming Institute, University of Nevada, Las Vegas, Las Vegas, NV 89154, USA
[5]Lee Business School, University of Nevada, Las Vegas, Las Vegas, NV 89154, USA

Corresponding author: Ismael Gómez-Talal (ismael.gomez.talal@urjc.es)

This work was supported in part by the Autonomous Community of Madrid (European Laboratory for Learning and Intelligent Systems (ELLIS) Madrid Node), and in part by the Spanish Ministry of Science and Innovation (LATENTIA) under Project PID2022-140786NB-C32/AEI/10.13039/501100011033.

ABSTRACT The phenomenon of cancellations in hotel bookings is one of the main pain points in the hospitality sector as it skews demand signals and can result in revenue losses estimated at about 20 %. Yet, forecasting booking cancellations remains an underresearched area, particularly in the understanding of the behavioral drivers of cancellations. This paper addresses this gap by proposing a new approach to predicting hotel booking cancellations rooted in stacked generalization and Explainable Artificial Intelligence (XAI). Specifically, the combination of linear, tree-based, non-linear and deep learning models into a single meta-model resulted in an increased accuracy rate to 96 %. In addition, this work focuses on interpretability, identifying the driving behavioral factors of cancellation as location, type of room, and customer segments. This approach can provide hoteliers with both highly accurate predictions as well as marketing intelligence that would allow them to drive strategy to minimize loss resulting from cancellations. The results of the research provide an effective solution to the challenges involved in forecasting booking cancellations, balancing forecast prediction accuracy with the ability to provide actionable insights.

INDEX TERMS Cancellation forecasting, hotel booking, artificial intelligence, machine learning, revenue management, explainable artificial intelligence.

I. INTRODUCTION

Tourism's meteoric rise as a cornerstone of global economic development is well documented, with the World Tourism Organization (UNWTO) indicating that the international tourism industry has recovered significantly in 2023, reaching approximately 1.3 billion international arrivals globally, about 88% of pre-pandemic levels. This strong recovery is expected to continue, with forecasts suggesting that international tourism could surpass pre-pandemic levels by the end of 2024 [1]. Aside from the income generated from tourists' direct consumption, the tourism sector also plays an important role in shaping infrastructural and service-related developments. Yet due to externalities like geopolitical flux, environmental conditions, and unforeseen calamities, the tourism is subject to a certain degree of unpredictability [2]. Nowhere is the brunt of these uncertainties felt more acutely than in the hospitality industry, where the perishable nature of the service implies that any unoccupied room represents an irrevocable loss in potential revenue [3], [4]. The forecasting of hotel demand, thus, is not merely an exercise in prediction, but a vital tool for strategic resource allocation and operational planning. It informs decisions across a spectrum from tactical to strategic management, encompassing aspects like inventory management, marketing initiatives, and staffing [5]. Yet, this task is notoriously complex due to the dynamic and unpredictable patterns of customer behavior, particularly when

The associate editor coordinating the review of this manuscript and approving it for publication was Muammar Muhammad Kabir.

factoring in cancellations, which significantly distort demand signals [6].

The advent of the internet has added layers of complexity to this challenge. It has facilitated an ease of booking (and by extension, cancellation) that has reshaped customer reservation habits. Guests are now prone to book multiple options or cancel due to minor changes in their preferences or plans, leading to significant impacts on cancellation rates [7]. In fact, previous studies have found that between 10-30% of hotel bookings are affected by cancellations [7], [8], which can in some cases account for losses in revenue of up to 20% [9]. Despite the far-reaching implications of this trend, the domain of cancellation forecasting, particularly at the individual customer level, remains underexplored, with limited insights into the drivers behind cancellation behaviors or effective preventative measures [10].

The implementation of Artificial Intelligence (AI) and AI-powered devices has revolutionized the creation and provision of services in the hospitality industry. Most of the research in the field regarding AI has been centered around its interaction with employees [11], customers [12], and organizational operation [13]. This study seeks to fill this gap by proposing a state-of-the-art Machine Learning (ML) model to accurately and explainably predict hotel booking cancellations. Explainability is crucial for practitioners to understand and communicate the decisions that are made by black-box AI algorithms. In the specific case of hotel booking cancellation, it may allow hoteliers to gain insights into the underlying factors driving cancellations, giving them the tools to proactively manage their inventory.

Accordingly, this study seeks to make a substantive contribution to the hotel demand forecasting literature by introducing an empirical meta-modeling approach rooted in stacked generalization to predict hotel booking cancellations. Eschewing reliance on vast historical datasets, this study uses a minimal set of variables to achieve high predictive accuracy, thereby streamlining the model training process and enhancing its responsiveness to market dynamics. Apart from achieving high predictive accuracy, this study also focuses on the interpretability of the model through the analysis of SHapley Additive exPlanations (SHAP) values, which yield insights into the individual variables driving booking cancellations. This dual focus on performance and interpretability through the meta-model ensures robust, accurate, and transparent predictions, positioning this research as a vanguard in the application of AI in hotel RM.

The rest of this paper is organized as follows. In Section II, existing literature is reviewed and a few important developments are mentioned concerning the application of ML to the prediction of hotel booking cancellations and interpretability techniques. Section III presents the framework of the research, containing the characteristics of the datasets used, the preprocessing methods which were carried out and the construction of the stacking meta-model framework. In Section IV, the results of the research are discussed, where the accuracy of different models is assessed in terms of their predictive ability and where the interpretative analysis with the use of SHAP-based methods is also detailed. In Section V, the focus is particularly on the practical and ethical application of the findings in relation to revenue management (RM). Finally, Section VI wraps up the article recounting the major aspects raised in the article, discussing the limitations, and proposing areas for future research.

II. LITERATURE REVIEW

In an era where AI has revolutionized the hospitality industry [14], revenue managers continue to grapple with the complexities of customer order cancellation. Despite the ubiquity of ML applications in other industries, the hospitality industry is still quite nascent, particularly in the context of revenue management [15]. This literature review reveals a promising trajectory for the utilization of ML algorithms, which calls for a strategic pivot towards profit-centric metrics and the innovative application of meta-modeling. By aggregating a diverse set of predictive models and extracting interpretability via SHAP values, this research seeks to improve the accuracy and explainability of hotel booking cancellation prediction.

A. PREDICTIVE ANALYTICS FOR HOTEL BOOKING CANCELLATIONS

The research stream of hospitality booking cancellation is a subset of the broader field focused on customer churn prediction. The use of ML techniques to identify the customers who are most likely to discontinue their services has been widespread, with significant contributions from researchers across various disciplines [16]. The finance sector [17], telecommunications [18], and online services [19] have seen a surge in studies aimed at understanding this phenomenon. Despite this, research surrounding customer churn specific to hospitality industry has been largely underexplored [20].

Cancellation prediction techniques originated in the travel industry. Traditionally, cancellation prediction studies have focused on transportation lines, such as airlines [21] or trains [22]. Yet despite its importance, research into hotel cancellation prediction is still relatively nascent.

The first application of ML to hotel cancellation prediction was done by [23], who used advanced booking data from a hotel chain to predict the average cancellation rate using multiple models, including linear, non-linear, and tree-based models. [9] proposed a customer choice based mathematical model for purchase and cancellation probabilities under different assumptions of cancellation rates. [24] employed a variety of ML algorithms on advanced booking data from individual hotels to classify bookings as cancellations.

In more recent studies, more sophisticated techniques such as k-nearest neighbors [25], boosting models [26], [27], and deep learning algorithms [8] were introduced. Furthermore, this research has demonstrated the effectiveness of both random forest and XGBoost models in accurately identifying potential cancellations [28], [29]. According to [30], Artificial Neural Networking with Genetic Algorithm (GA) model as the most suitable for predicting booking cancellations. [8]. XGBoost, in particular, has been lauded for its high computational efficiency and accuracy, making significant strides in addressing not only hospitality-related issues but also broader economic and business challenges [31], [32]. These more complex algorithms have shown remarkable accuracy [33], [34], but they have been criticized for their lack of interpretability.

Interpretability of ML models is crucial in forecasting literature, as it allows researchers and practitioners to gain insight into the drivers of the phenomena being modeled [35]. Yet, there have been only two previous booking cancellation studies that incorporate interpretability measures into ML models. First, [28] introduced interpretability measures through the built-in feature importance metric that the random forest model provides. More recently, [26] incorporated SHAP values into their profit-driven XGBoost model. As evidenced by the shortage of research, deeper insight into interpretable ML techniques is sorely needed in the booking cancellation literature.

B. DRIVERS OF BOOKING CANCELLATIONS

The cancellation of a hotel booking can be affected by a diverse range of internal or external factors. Internal factors can be inherent to either the customer or the boooking, while external factors can be related to larger, macro-level trends that may affect cancellation behavior.

In relation to the customer, cancellations may be caused by the customer's behavior or characteristics. In terms of behavior, consumers may purposely cancel their booking because of a better offering they found at a different hotel, such as a more competitive price, a better location, or more amenities [10]. Customer characteristics that influence their likelihood to cancel their booking are the customer country of residence (origin), type, and size of travel group [36]. [10] also suggested that the customer's previous cancellation behavior is another cancellation driver. That said, customer-related variables can be affected by the cancellation policy of the hotel organization [33], as time leniency (length of cancellation window) and monetary leniency (cost of cancellation fee) affect customer cancellation behavior [37]. Similarly, they can be influenced by other factors that may be out of the customer's control, such as weather, illness, and changes of plans [36].

Characteristics of the booking itself also play a key role in uncovering cancellation behavior. These can be in the form of product characteristics such as room rate or room type [38]. They can also be variables that are associated with temporal features of the booking, like arrival date, length of stay, and lead time [39]. For example, bookings made well in advance and shorter lengths of stay or booking pace have higher probabilities of cancellation [36]. Other temporal effects related to the booking that may affect cancellations are time of year (seasonality) as well as day of the week [40]. The distribution channel used by the client also affects cancellation behavior [20], where bookings from travel agencies were most likely to keep their booking.

External factors that influence booking cancellations are related to the environment within which the hotel is situated, such as competition, catastophe, macroeconomic performance, weather, and social reputation [10]. Competition from other hotels in the area can lure customers away with special rates or more amenities [41]. Catastophic events such as the COVID-19 or SARS outbreaks can lead to mass cancellations and may not be accounted for in forecasting models [42]. Macroeconomic performance has also been found to cause cancellations, not just for the hospitality industry, but also for the travel, events, sports, and entertainment industries [43]. Weather data has also been found to influence cancellation behavior as well as improve the prediction of ML models [44]. Lastly, [45] found that social reputation has a significant effect on hotel booking cancellation behavior, particularly when it comes to online reviews.

C. EVALUATING PREDICTIVE MODELS FOR HOSPITALITY BOOKING CANCELLATIONS

Optimizing predictive models for order cancellations involves a thorough comparison of diverse models, typically using statistical performance metrics. This conventional method has been criticized for its lack of focus on the core business aim of profit maximization through customer retention [46]. Therefore, profit-centric metrics such as the maximum profit criterion have been suggested to align model evaluations more closely with business goals, proving to significantly influence retention campaign profitability [47]. Accordingly, this study proposes the adoption of these profit-based metrics over traditional statistics to enhance the prediction of order cancellations, filling the void left by prior studies (see Table 1).

D. ADVANCING MODEL SELECTION THROUGH META-MODELING

The meta-model, often referred to as a 'stacking' model, is a two layer model, where a first layer of several base learners is trained, and the second layer integrates the predictions of the base learners using another model to consolidate and fine-tune their predictions [48]. By combining multiple models, the meta-model aims to optimally combine the prediction results of the base learners to gain predictive power from the ensemble of their predictions and mitigate their individual weaknesses [49]. This technique is particularly pertinent when models exhibit a wide variability in performance across different datasets and validation methods, as indicated in the summary of related works (see Table 1).

In contrast to the individual models tested in previous studies, the meta-model capitalizes on the diversity of the diverse range of the forecasting models that exist in the literature. It dynamically incorporates the heterogeneous

TABLE 1. Summary of validation and hyperparameter selection methods in related works, alongside the interpretability analysis and the best model selected.

Reference	Validation	Hyperparameters Selection	Interpretability	Best Model Selected
[20]	10-fold CV	—	—	Boosted Decision Tree
[7]	10-fold CV	Specific tuning	—	XGBoost
[52]	10-fold CV	—	—	XGBoost
[8]	Random sampling	Genetic algorithm	—	Artificial Neural Networking
[28]	10-fold CV	Fine-tuning	Feature importance	Random Forest
[34]	—	Grid search	—	Random Forest
[53]	5-fold CV	Grid search	—	Lasso-Bayesian
[26]	10-fold CV	Grid search	SHAP	Profit-Driven XGBoost
This study*	5-fold CV	Grid search	SHAP	Meta-Model

*Note: Validation metrics in this table may be computed using the bootstrap method, which involves random sampling with replacement to estimate the sampling distribution and to provide more robust metrics.

perspectives offered by each approach, giving it the potential to yield a more comprehensive and nuanced understanding of customer behavior. This is backed empirically by the improvements in predictive power that stacking has shown in comparison to individual base learners [50], [51]. The meta-model efficacy is further bolstered by the application of SHAP values for interpretability, which providing insights into the feature contributions across all stacked models. As such, the meta-model stands as a significant advancement in model selection for the hospitality industry.

E. INTERPRETABLE ML

As AI permeates through all sectors of the hospitality industry, the need for interpretable ML models have become increasingly important. Historically, one of the concerns with respect to AI and ML methodologies has been the trade-off between interpretability and predictive accuracy [54]. ML models provide high predictive power, but are often black box algorithms that give the user no information about what is actually driving their outputs. This lack of transparency can lead to discrimination, distrust, secrecy, and unethical or illegal practices [55], [56]. To mitigate these risks, researchers have created techniques to make models more interpretable, or understable by human users [57].

1) IMPORTANCE OF INTERPRETABILITY

Interpretability in ML is essential for several reasons: it enhances transparency, aids in compliance with legal standards, and facilitates error identification and model improvement [58]. In the context of hotel booking cancellations, interpretable models help stakeholders understand the rationale behind AI-driven decisions, such as the prediction that a particular customer may cancel their booking [57], [59]. This will allow the hotelier to understand the underlying trends driving cancellations without sacrificing prediction accuracy. This deeper understanding of booking cancellation trends may consequently aid in revenue management functions such as demand forecasting, inventory management, and pricing strategy.

2) ETHICAL CONSIDERATIONS IN AI

Ethical AI encompasses issues such as fairness, privacy, accountability, and transparency [60]. In the hospitality industry, these issues translate into ensuring that AI models do not inadvertently discriminate against certain customer groups, are capable of explaining their decisions when required, and allow for human oversight where necessary [61]. For instance, ensuring that a cancellation prediction model does not disproportionately target guests from specific regions or demographic backgrounds is vital for maintaining fairness.

3) LITERATURE ON INTERPRETABLE ML

Significant strides have been made in developing interpretable ML techniques. Methods such as SHAP and Local Interpretable Model-agnostic Explanations (LIME) have been particularly prominent in providing insights into the contributions of individual features in model predictions [62]. Yet, significant challenges remain in fully integrating ethical AI practices within the hospitality industry [14]. These include balancing the trade-off between model complexity and interpretability, privacy and data protection, legal and human rights, and the ongoing need to update ethical guidelines as AI technologies evolve [63], [64]. However, these challenges also present opportunities for advancing research in AI applications that are not only technically proficient, but also ethically sound.

III. METHODOLOGY

This section delves into the comprehensive methodology adopted for predicting booking cancellations in the hospitality sector. Our approach meticulously pre-processes a rich dataset, applies a variety of sophisticated ML models, and integrates these models into a meta-model using a stacking technique. The entire process, from preprocessing to final evaluation, is visualized in a research diagram, providing a clear and structured overview of the steps undertaken in this study.

A. DATA AND PREPROCESSING

The data utilized in this study was generously provided by a Southern European hotel with an official classification of four stars. The dataset contains 79,330 records of booking data for a period of two years. A comprehensive summary of the variables included in the dataset is detailed in Table 2.

This dataset encapsulates a range of factors including, but not limited to, customer demographics, stay details, and booking information. Each variable was meticulously selected for its relevance in developing a robust ML model capable of predicting the cancellation of a booking with high accuracy.

One-hot encoding is a popular technique for preprocessing categorical data in ML. This method converts categorical variables, which are often represented as a string of text or a single-label encoding, into a binary vector representation that can be more effectively utilized by ML algorithms. When employing one-hot encoding, each category level of the variable becomes a new attribute in a binary format, where only one attribute will be 'hot' (i.e., set to 1) while all others are 'cold' (set to 0) for each observation [65].

This transformation is vital as it allows algorithms to exploit the presence or absence of a category without imposing ordinality where it may not exist [66]. For example, assigning numerical values like 1, 2, 3 to categories such as 'small', 'medium', 'large' can be misleading to an algorithm, as it may infer that 'large' is three times more significant than 'small', which is semantically incorrect and can lead to biased results. One-hot encoding avoids this by treating each category as an independent feature, thus preserving the categorical information without implying any ordinal relationship [67].

Furthermore, one-hot encoding is often used to overcome the limitation of models that are unable to work with categorical data directly, such as certain regression models, support vector machines, and neural networks [68]. By converting categories into numerical arrays, one-hot encoding enables the use of categorical data within these models, thereby enhancing their applicability across a wider range of datasets [69].

To prepare the dataset for model training, the following set of categorical variables were one-hot encoded: *Meal*, *Country*, *MarketSegment*, *DistributionChannel*, *ReservedRoomType*, *AssignedRoomType*, *DepositType*, and *CustomerType*. The one-hot ransformation ensured that no ordinal relationships were imposed into categorical variables which would bias the model. Such features enabled the models to have higher accuracy rates since the predictors such as booking segment and meal type were highly correlated with booking cancellations.

B. BINARY CLASSIFICATION MODELS AND STACKING METAMODEL APPROACH

We tested eight binary classification models across various combinations of three distinct math score levels: low, medium, and high. Each model was trained using the Stratified K-fold method, ensuring data balance within each fold through the application of an undersampling technique. For optimization, every model employed differing hyperparameters, followed by the introduction of a grid search to pinpoint the model that delivered the highest performance. The evaluated models were the following.

Logistic Regression (LR): LR is a statistical method for predicting binary classes. This means that the outcome or target variable is dichotomous in nature (e.g., 'yes' or 'no'). LR estimates the probabilities using a logistic function, which is an S-shaped curve that can take any real-valued input and map it between 0 and 1. The equation of the logistic regression is:

$$\sigma(z) = \frac{1}{1 + e^{-z}}, \tag{1}$$

where $\mathbf{z}$ is the linear combination of input features $\mathbf{x}$ and their corresponding weights $\mathbf{w}$, given by $\mathbf{z} = w_0 + w_1x_1 + w_2x_2 + \ldots + w_nx_n$. LR becomes a binary classifier by setting a threshold, typically 0.5, to decide the predicted class.

Decision Trees (DT): A decision tree is a flowchart-like structure where each internal node denotes a test on an attribute, each branch represents the test outcome, and each leaf node corresponds to a class label (in classification) or a continuous value (in regression). Decision trees classify instances by traversing from the root to a leaf node, which provides the classification. Criteria such as entropy or Gini impurity determine splits at internal nodes [70].

Entropy measures the impurity in a set, defined for a set $\mathbf{S}$ as:

$$Entropy(\mathbf{S}) = -\sum_{i=1}^{c} p_i \log_2(p_i), \tag{2}$$

where c is the number of classes, and p_i is the proportion of examples in $\mathbf{S}$ that belong to class i.

Gini impurity measures the frequency of incorrect classification, defined for a set $\mathbf{S}$ as:

$$Gini(\mathbf{S}) = 1 - \sum_{i=1}^{c} p_i^2. \tag{3}$$

Random Forest (RF): Random Forests improve upon decision trees by creating an ensemble of trees to improve accuracy and control overfitting. They introduce randomness through bootstrap aggregating and feature randomness [71].

Given a training data matrix $\mathbf{X}$ and class labels $\mathbf{Y}$, a random forest classifier RF is defined by the mode of classifications from its trees $T_1, T_2, \ldots, T_n$ for an instance $\mathbf{x}$:

$$RF(x) = \text{mode}\{T_1(\mathbf{x}), T_2(\mathbf{x}), \ldots, T_n(\mathbf{x})\}, \tag{4}$$

where T_i is i-th decision tree in the forest, and $\mathbf{x}$ is instance being classified.

Gradient Boosting (GB): GB is a ML technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees [72]. It builds the model in a stage-wise fashion like other boosting methods do, and it generalizes them by allowing optimization of an arbitrary differentiable loss function. In the context of binary classification, gradient boosting modifies its model by sequentially adding new models that correct the errors made

TABLE 2. Summary of dataset variables for hotel booking cancellation prediction. The dataset comprises various factors related to hotel stays, customer characteristics, and booking details. Each variable is crucial for developing a predictive model to determine the likelihood of booking cancellations.

Variable	Description	Type
LeadTime	Number of days between the booking date and the arrival date.	Continuous
ArrivalDateYear	Year of arrival.	Categorical (Ordinal)
ArrivalDateMonth	Month of arrival.	Categorical (Ordinal)
ArrivalDateWeekNumber	Week number of the year for arrival.	Continuous
ArrivalDateDayOfMonth	Day of the month of arrival.	Continuous
StaysInWeekendNights	Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel.	Continuous
StaysInWeekNights	Number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel.	Continuous
Adults	Number of adults.	Continuous
Children	Number of children.	Continuous
Babies	Number of babies.	Continuous
Meal	Type of meal booked. Categories are presented in standard hospitality meal packages.	Categorical (One-Hot Encoded)
Country	Country of origin of the guests.	Categorical (One-Hot Encoded)
MarketSegment	Market segment designation. In categories such as "Online TA", "Offline TA/TO", "Groups", etc.	Categorical (One-Hot Encoded)
DistributionChannel	Booking distribution channel. The term TA/TO means "Travel Agents/ Tour Operators".	Categorical (One-Hot Encoded)
IsRepeatedGuest	Indicates if the guest is a repeat customer (1) or not (0).	Binary
PreviousCancellations	Number of previous bookings that were canceled by the customer prior to the current booking.	Continuous
PreviousBookingsNotCanceled	Number of previous bookings not canceled by the customer prior to the current booking.	Continuous
ReservedRoomType	Code of room type reserved. Room types are coded by different letters.	Categorical (One-Hot Encoded)
AssignedRoomType	Code for the type of room assigned to the booking. Sometimes the assigned room type differs from the reserved room type.	Categorical (One-Hot Encoded)
BookingChanges	Number of changes/amendments made to the booking from the moment the booking was entered onto the system until check-in.	Continuous
DepositType	Indicates if the customer made a deposit to guarantee the booking. Categories include "No Deposit," "Non-Refund," and "Refundable".	Categorical (One-Hot Encoded)
DaysInWaitingList	Number of days the booking was on the waiting list before it was confirmed to the customer.	Continuous
CustomerType	Type of booking, assuming one of four categories.	Categorical (One-Hot Encoded)
ADR	Average Daily Rate, calculated by dividing the sum of all lodging transactions by the total number of staying nights.	Continuous (Normalized)
RequiredCarParkingSpaces	Number of car parking spaces required by the customer.	Continuous
TotalOfSpecialRequests	Number of special requests made by the customer.	Continuous
IsCanceled	Indicates if the booking was canceled (1) or not (0).	Binary

by existing ensemble models using a gradient descent like procedure:

$$GB(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^{n} \alpha_i T_i(\mathbf{x})\right), \tag{5}$$

where T_i represents the i-th decision tree, α_i is the weight of the i-th tree, and $\mathbf{x}$ is the input feature vector.

Support Vector Machine (SVM): SVM is a powerful classifier that works both in linear and non-linear cases. For binary classification, SVM aims to find the best margin (distance between the line and the support vectors) that separates the two classes in the feature space. This is achieved by solving an optimization problem that maximizes the margin between the classes' closest points, known as support vectors. The decision function is defined as:

$$SVM(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^{n} \alpha_i y_i \langle \mathbf{x}, x_i \rangle + b\right), \tag{6}$$

where $\mathbf{x}_i$ are the support vectors, $\mathbf{y}_i$ their corresponding labels, α_i are the coefficients, and b is the bias.

XGBoost. XGBoost stands for Extreme Gradient Boosting, an efficient and scalable implementation of gradient boosting. It provides a parallel tree boosting that solves many data science problems in a fast and accurate way. XGBoost improves upon the base GBM (Gradient Boosting Machine) framework through systems optimization and algorithmic enhancements, such as handling missing data, regularizing to prevent overfitting, and employing a more efficient tree learning algorithm. XGBoost specifically defines an objective function as a combination of a specific loss function and a regularization term:

$$XG(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^{n} T_i(\mathbf{x})\right), \tag{7}$$

where $T_i(\mathbf{x})$ represents the output of the i-th boosted tree for the instance $\mathbf{x}$.

LightGBM: A Gradient Boosting Decision Tree (GBDT) model, addresses the efficiency and scalability challenges faced by traditional GBDT implementations when dealing with high-dimensional data and large datasets. Two novel techniques, Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB), are introduced to enhance the model performance without compromising accuracy. GOSS focuses on retaining data instances with large gradients while performing random sampling on instances with small gradients. This approach is grounded in the observation that instances with larger gradients contribute more significantly to the information gain, which is a crucial factor in determining the best split points in decision trees.

Given a dataset with instances $\{x_1, x_2, \ldots, x_n\}$ and their corresponding gradients $\{g_1, g_2, \ldots, g_n\}$, GOSS proceeds as follows: (1) sort the instances by the absolute values of their gradients in descending order; (2) select the top $a \times 100\%$ instances with the largest gradients; (3) randomly sample $b \times 100\%$ of the remaining instances; and (4) for the split point evaluation, amplify the contribution of the randomly sampled instances by a constant factor to preserve the data distribution.

The effectiveness of GOSS is demonstrated through theoretical analysis, showing that it achieves a more accurate information gain estimation compared to uniform random sampling, especially when the range of information gain values is large.

EFB aims to reduce the number of features in sparse datasets by bundling mutually exclusive features, that is, features that rarely take nonzero values simultaneously [73]. This reduction is nearly lossless and significantly decreases the computational complexity of building feature histograms, a critical step in GBDT training [74].

The key steps in EFB include: (1) construct a feature graph where vertices represent features, and edges represent the non-exclusivity between features; (2) apply a greedy algorithm to color the graph, where each color corresponds to a bundle of exclusive features; and (3) merge the exclusive features into bundles, ensuring that the original feature values can be recovered from the bundled feature values.

EFB effectively reduces the feature space dimensionality, enhancing the training speed of the LightGBM model without sacrificing model accuracy [75].

Multilayer Perceptron (MLP): MLP is a class of feedforward artificial neural network (ANN). An MLP consists of at least three layers of nodes: an input layer, a hidden layer, and an output layer. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. MLP utilizes a supervised learning technique called backpropagation for training. Its multiple layers and non-linear activation distinguish MLP from a linear perceptron. It can distinguish data that is not linearly separable:

$$MLP(\mathbf{x}) = f(W_n \cdot f(W_{n-1} \cdots f(W_1 \cdot x + b_1) \cdots) + b_n), \tag{8}$$

where f is the activation function, $\mathbf{W}_i$ and $\mathbf{b}_i$ are the weight matrix and bias vector for the i-th layer, respectively, and $\mathbf{x}$ is the input feature vector. The function f represents the activation function applied at each layer, transforming the linear combination of inputs and weights to produce a non-linear output, essential for learning complex patterns in the data.

1) STACKING METAMODEL APPROACH FOR HOTEL CANCELLATION

The technique of stacking, initially introduced by [48], represents a significant leap forward in the ML domain. It suggests an ensemble method that leverages the output of several models to train a superior meta-model. This strategy is instrumental in reducing generalization error, mitigating bias, and enhancing prediction results [76]. Meta-stacking enhances the basic idea of stacking by iteratively merging base learners with meta-models, thus underlining the perpetual significance of Wolpert's contribution to the field of ML.

Upon reviewing eight distinct binary classification models for hotel cancellation prediction, we proceed to apply a stacking metamodel strategy. The meta-learner amalgamates the predictions of the eight models into one cohesive output, heightening the accuracy and robustness of the forecast results [48], [77].

Initially, an optimal individual model of each base algorithm type was determined through a deep hyperparameter search that was injected into the stacking approach. To evaluate each model fairly and robustly, the hyperparameters of each model were fine-tuned via the grid search with stratified k-fold cross validation. After the previous processes, the model outputs created from the trained dataset were brought together to form the stacking meta-model. The goal of the stacking framework was to combine the benefits of these optimized base models.

a: EXHAUSTIVE HYPERPARAMETER SEARCH FOR BASE MODELS

Grid search is the most commonly-used method in recent cancellation prediction literature [26], [53]. A grid search involves systematically evaluating different combinations of hyperparameters by exhaustively searching through a predefined grid [78]. A grid search was conducted for all eight of the ML models deployed in this study. The goal of this exhaustive search is to independently optimize the performance of each model, ensuring that only the top-performing versions are considered for integration into the stacking framework.

b: SELECTION AND TRAINING OF THE META-MODEL

Following the optimization of base models, we incorporate their prediction outputs as the various inputs for the meta-model. For this analysis, LR was chosen as the meta-classifier, as it is effective for meta-model classification problems [79]. This is because of its proven effectiveness in binary classification tasks, straightforwardness, interpretability, and the linear characteristic of its decision boundary [80]. The training regimen for the meta-model also entails a hyperparameter optimization phase, tailored specifically to LR. This stage adjusts the LR model to optimally assimilate the inputs from the base models, ensuring that the stacking meta-model leverages the combined strengths of the individual models to the fullest.

c: IMPLEMENTATION DETAILS

Our stacking framework, spanning from base models to the meta-model, employs a stratified K-fold cross-validation method. This strategy effectively preserves the distribution of target variable instances across folds, allowing the model to better generalize to unseen observations [81]. The comprehensive search for hyperparameters and the subsequent LR meta-model training are conducted with an acute awareness of overfitting risks and computational efficiency considerations.

d: EVALUATION AND OPTIMIZATION

The performance of the various ML models were assessed using a broad set of metrics. The focal point of the optimization endeavors is the hyperparameters of the LR meta-model, following the thorough search and refinement of base models. This bottom-up strategy ensures that the stacking method not only harnesses the predictive power of individual models but also efficiently consolidates their predictions to maximize prediction accuracy. Although voting is frequently employed in ensemble approaches, it is not the same as the stacking approach used in this research. Voting is a method of combining predictions without having to train a new model. Random Forest and Gradient Boosting Decision Trees employ bagging and boosting to form strong models from weak learners. On the other hand, stacking builds a meta-model that combines the outputs of the diverse base models integrated into it. This method effectively utilizes the advantages of the base models instead of average prediction which enhances out-of-sample accuracy and robustness of the ensemble method.

C. STRATIFIED K-FOLD CROSS-VALIDATION FOR PREDICTIVE MODEL TRAINING

Stratified K-Fold cross-validation is especially beneficial for datasets with imbalanced classes, such as hotel cancellations [25]. This method ensures an even distribution of class instances across all folds, facilitating a fair and consistent evaluation of model performance [82].

In this study, the training set (accounting for 80% of the total dataset) was organized by the target class of cancellation and divided into k = 5 stratified folds, such that each fold reflected the class proportions of the entire dataset accurately. Throughout each validation cycle, one fold served as the validation set while the remaining folds constituted the training set. The model was then trained on this aggregated set and evaluated against the validation set [83].

This process was repeated five times, treating each fold as a validation set once. It is crucial to note that the models were recalibrated in each fold of the cross validation process to avoid any influence from previously fitted models [84]. Finally, the performance measures for each fold were averaged to obtain a robust and accurate assessment of model effectiveness [85].

D. MODEL EVALUATION

The accuracy metrics used in this study are: Accuracy, Recall (RC), F1 Score, Precision (PR), Specificity (SP), and Area Under the Curve (AUC) [86]. In the case of the first metric, it provides an overall measure of model performance by indicating the proportion of correct predictions compared to the total number of predictions made. The accuracy is calculated as

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{9}$$

where the TP value represents the true positive rate, the TN value represents the true negative rate, the FP value represents the false positive rate and the FN value represents the false negative rate.

The RC value represents the model ability to correctly identify all positive cases, being especially useful in situations where FN are more problematic than FP. It can be mathematically represented as

$$RC = \frac{TP}{TP + FN}. \tag{10}$$

PR measures the proportion of correctly classified positive samples (true positives) among all samples classified as positive ($TP + FP$), defined as

$$PR = \frac{TP}{TP + FP}. \tag{11}$$

SP, on the other hand, measures the proportion of correctly classified negative samples (true negatives) among all samples classified as negative (true negatives + false positives) as defined by the following equation

$$SP = \frac{TN}{TN + FP}. \tag{12}$$

The F1 Score is considered a balanced measure between accuracy and completeness. It is calculated as follows:

$$F1Score = \frac{2 \times (PR \times RC)}{PR + RC}. \tag{13}$$

Finally, the AUC, which represents the area under the receiver operating characteristic (ROC) curve, is used to evaluate the totality of two parameters: the true positive rate and the false positive rate. Where the equation is defined as the integral that calculates the area as follows

$$AUC = \int_0^1 \mathrm{ROC(t)dt}, \tag{14}$$

where an AUC of 1 denotes a perfect classification model, while a value of 0.5 suggests a performance no better than random classification [87], [88].

E. SHAPLEY VALUES FOR INTERPRETABLE MACHINE LEARNING

The term Shapley value came from game theory, which interpreted features of a model as players and the task of making the prediction as a cooperative game [89], [90]. In AI, they describe the role of single features with a model output and give an attribution of the prediction value across the features' [91]. More specifically, the Shapley value corresponds to the deviation of the prediction at a particular query point from the average prediction due to the feature in question. For each query point, the sum of the Shapley values for all features corresponds to the total deviation of the prediction from the average. Mathematically, the Shapley value of the i-th feature for query point x is defined by the value function v_x as

$$\varphi_i(v_x) = \frac{1}{M} \sum_{S \subseteq M_s\{i\}} \frac{v_x(S \cup \{i\}) - v_x(S)}{(M-1)!}, \tag{15}$$

where M is the total number of features, M_s is the set of all features, $|S|$ is the cardinality of the set S, i.e., the number of elements in the set S and $v_x(S)$ is the value function of the features in a set S for query point x (indicates the expected contribution of the features in S to the prediction for query point x).

Several algorithms are provided to compute these values within the SHAP package in Python. These algorithms, often called *interventional algorithms*, include variants such as Kernel SHAP, Linear SHAP, and Tree SHAP.

SHAP Intervention Algorithms Intervention algorithms define the value function for a set of attributes at a query point as the expected prediction for the intervention distribution, which is the joint distribution of the attributes in the complement of the set

$$v_x(S) = E_D[f(x_S, X_{S^c})], \tag{16}$$

where x_S is the query point value for the features in S, and X_{S^c} are the features in S^c. The intervention algorithm evaluates the value function $v_x(S)$ at query point x, under the assumption that the features are not highly correlated, using the values in the data X as samples from the intervention distribution D for the features in S^c

$$v_x(S) = E_D\left[f\left(x_S, X_{S^c}\right)\right] \approx \frac{1}{N}\sum_{j=1}^{N} f\left(x_S, (X_{S^c})_j\right), \tag{17}$$

where N is the number of observations, and $(X_{S^c})_j$ contains the values of the features in S^c for the j-th observation.

The main advantage of interventional algorithms is that they are computationally efficient, although they do require the assumption of feature independence and use samples outside the distribution [92]. This could potentially result in unrealistic observations [93]. The main SHAP intervention algorithms include the following:

Kernel SHAP is a version of the Shapley algorithm that uses a kernel approximation to estimate the Shapley values. This method is particularly useful when faced with a high-dimensional feature space but can be computationally intensive [62].

Linear SHAP is used when we have a linear model, and it calculates the Shapley values analytically rather than approximately. This makes Linear SHAP computationally more efficient than Kernel SHAP [62].

Tree SHAP is a variant of the Shapley algorithm designed specifically for tree-based models, such as decision trees, random forests, and GB algorithms. Tree SHAP is particularly computationally efficient and can handle feature interactions explicitly [94], [95].

F. RESEARCH DIAGRAM

The research diagram presented in Figure 1 provides a visual representation of the methodological framework employed for the prediction of booking cancellations in this study. First, the dataset was preprocessed to ensure quality and consistency. Then, a stratified 5-fold cross-validation was employed to maintain the proportion of each class within each fold.

The methodology encompasses the training of various predictive models: LR, SVM, DT, RF, GB, XGBoost, LightGBM, and MLP. These models are carefully chosen to cover a broad spectrum of ML techniques, from simple linear models to more complex ensemble and neural network approaches. This is to ensure model diversity, which is a focus of modern forecast aggregation methodology [96], as it can improve meta-model performance [76], [97]. Upon training, the individual models are integrated into the stacking ensemble. The performance of the ensemble is then evaluated through a comprehensive set of accuracy metrics. These metrics provide a holistic view of the models' predictive performances, from overall accuracy to the ability to balance false positives and false negatives.

Interpretability is also a key consideration in our framework. The best-performing meta-model is subjected to an interpretability analysis using SHAP values. This step ensures that the predictions of the model are not only accurate but also understandable and explainable. Explainability is important for practical applications where stakeholders may require insights into the model decision-making process, as well as to soothe ethical concerns regarding black-box models.

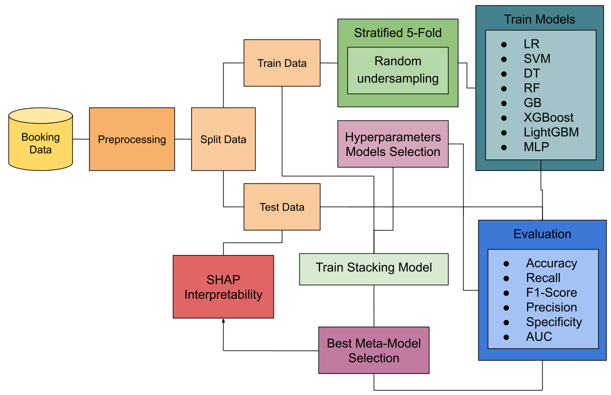


FIGURE 1. Research methodology for predicting booking cancellations, including data preprocessing, model training with stacking ensemble, and evaluation using SHAP values and key performance metrics.

IV. RESULTS

This section details the results of our comprehensive ML framework established to evaluate the performance of various ML models in predicting hotel booking cancellations. First, a comparative analysis of the forecasting models is discussed. Then, the interpretability analysis of the forecasting models is presented. These empirical analyses are designed to not only assess the predictive power of individual models, but also to investigate the enhanced performance and generalizability achievable through the integration of a meta-model. The overarching goal is to identify a predictive framework that combines high accuracy with the interpretability essential for practical and ethical application in the hospitality industry.

A. COMPARATIVE ANALYSIS OF FORECASTING MODELS

1) DESCRIPTION OF EXPERIMENTS

This subsection offers a complete description of the experiments conducted in the research. Table 4 presents the full set of experiments performed, evaluations of the eight machine learning models, and several of their variants developed as a result of varying the hyperparameters tuning process. It also includes the meta-model that was constructed by stacking, which aims at obtaining better prediction results by combining the advantages of the individual models. textbfModel Selection. Table 3 presents a detailed comparative analysis of different hyperparameter tunings for the eight booking cancellation models and the meta-model. The models were evaluated using the following metrics: ACC, RC, F1 Score, PR, SP, and AUC.

2) PERFORMANCE SUMMARY

Table 3 succinctly summarizes the performance of binary classification models in predicting hotel booking cancellations. Taking the best hyperparemeter tuning for each model, it compares the predictive performance of the eight models across the six accuracy metrics. Among the base models, RF was the best performer with an accuracy of 93.5%. The weakest base learner was the LR model, which yielded an accuracy of only 80%. While the base learners alone yielded high prediction accuracies, they were overshadowed by the meta-model across all accuracy metrics. These results underscore the efficacy of the stacking model in leveraging the combined strengths of the individual models to accurately forecast cancellations.

B. INTERPRETABLE ANALYSIS OF HOTEL BOOKING CANCELLATIONS

The SHAP value analysis, as presented in Figure 2, offers a comprehensive overview of the features that contribute to the model predictions.

Figure 2 (a) depicts the mean SHAP values, which quantify the average impact of each variable on the model output. These values provide insight into the general importance of each feature across all predictions. For instance, variables such as the country of origin, lead time before arrival, and room type are shown to have substantial influence on the likelihood of cancellation.

Figure 2 (b) extends the analysis by displaying the distribution of SHAP values for each feature, shedding light on how individual feature levels affect the prediction. Positive SHAP values suggest an increased probability of booking cancellation, while negative values imply a decreased probability. This subfigure aids in understanding the heterogeneity of effects across different feature values.

Lastly, Figure 2 (c) presents a heatmap of SHAP values across individual instances tested. This detailed visualization enables the identification of patterns and anomalies in the data, showing the direction (positive or negative) and magnitude of each feature impact on specific instances.

The heatmap generated from SHAP values is a powerful visualization that assists in understanding how different features influence a predictive model decision on whether a hotel booking will be canceled or not. The color intensity represents the impact of each feature, with red indicating a higher impact and blue a lower impact on the model output. The SHAP values on the right side of the heatmap correlate with an increased likelihood of cancellation, whereas the values on the left side correlate with a decreased likelihood. Therefore, features represented by stronger red colors on the right contribute to predicting a higher chance of cancellation, and those with stronger blue colors on the left contribute to predicting a lower chance of cancellation.

When explaining the importance of the variables:

1) HIGH POSITIVE SHAP VALUE (RED)

Features that frequently show a strong red color, especially toward the right side of the heatmap, are crucial in pushing the model towards predicting a cancellation. These variables may represent characteristics that are common among bookings that ended up being canceled.

2) HIGH NEGATIVE SHAP VALUE (BLUE)

Conversely, features that appear with a strong blue color on the left are significant in influencing the model to predict a non-cancellation. These features might be indicators of

TABLE 3. Performance comparison of binary classification models for predicting hotel booking cancellations. The table showcases the effectiveness of each model, underscored by the introduction of a meta-model that leverages the combined strengths of individual models.

Model	Accuracy	Recall	F1 Score	Precision	Specificity	AUC
LR	0.8081 ± 0.0025	0.7579 ± 0.0054	0.7668 ± 0.0033	0.7760 ± 0.0043	0.8439 ± 0.0042	0.8928 ± 0.0021
SVM	0.8450 ± 0.0254	0.9009 ± 0.0310	0.8168 ± 0.0277	0.7478 ± 0.0343	0.8102 ± 0.0325	0.9091 ± 0.0106
DT	0.9114 ± 0.0019	0.9205 ± 0.0043	0.8964 ± 0.0021	0.8735 ± 0.0041	0.9049 ± 0.0038	0.9158 ± 0.0022
RF	0.9354 ± 0.0018	0.9231 ± 0.0035	0.9224 ± 0.0021	0.9218 ± 0.0034	0.9442 ± 0.0026	0.9834 ± 0.0008
GB	0.8419 ± 0.0025	0.8057 ± 0.0050	0.8093 ± 0.0030	0.8129 ± 0.0047	0.8677 ± 0.0044	0.9262 ± 0.0020
XGBoost	0.8676 ± 0.0032	0.8438 ± 0.0050	0.8414 ± 0.0035	0.8392 ± 0.0043	0.8847 ± 0.0038	0.9465 ± 0.0017
LightGBM	0.8778 ± 0.0022	0.8597 ± 0.0042	0.8541 ± 0.0026	0.8487 ± 0.0041	0.8907 ± 0.0034	0.9534 ± 0.0015
MLP	0.8878 ± 0.0054	0.8935 ± 0.0115	0.8689 ± 0.0057	0.8460 ± 0.0138	0.8837 ± 0.0130	0.9566 ± 0.0021
Meta-Model: LR	$\mathbf{0.9689 \pm 0.0010}$	$\mathbf{0.9620 \pm 0.0021}$	$\mathbf{0.9761 \pm 0.0008}$	$\mathbf{0.9906 \pm 0.0011}$	$\mathbf{0.9823 \pm 0.0021}$	$\mathbf{0.9951 \pm 0.0065}$

bookings that are more likely to be fulfilled rather than canceled.

3) FEATURES CLOSE TO ZERO (WHITE)

Features that have colors close to white, indicating SHAP values near zero, have little to no impact on the prediction in either direction.

Among the variables, the origin of the booking emerges as a significant predictor. When bookings originate from Portugal (‘Country_PRT’), there is a noticeable increase in the likelihood of cancellation, as indicated by the prominent red shading associated with this feature. This suggests that local bookings may have a higher propensity to be canceled, possibly due to the ease of changing plans. Foreign booking show different results. Bookings that originate in Italy (‘Country_ITA’) show high positive SHAP values, yet bookings from Spain (‘Country_ESP’) show high negative SHAP values. This means that bookings from Italy have higher probabilities of cancellation, while bookings from Spain have lower probabilities of cancellaiton. This could be due to physical proximity, market-specific behavior, or economic factors in each country. For example, the average Spanish customer is much closer to the hotel than the average Italian customer.

‘LeadTime’, or the number of days between the booking and the intended stay, is another critical variable. A longer lead time strongly correlates with cancellations, evidenced by its red intensity on the right. This could imply that customers who plan far ahead are more susceptible to unforeseen changes, leading to cancellations.

Conversely, ‘AssignedRoomType_A’ and ‘AssignedRoomType_D’ show deep blue sections on the left side of the heatmap, indicating that guests in those room types were less likely to cancel. The room types ‘A’ and ‘D’ were the rooms with the lowest ADR and the most inventory. These results indicate that customer cancellation behavior varies by room type. The assignment of a specific room type might reflect a firm commitment from the hotel, which can give customers greater confidence in their booking.

The variable ‘BookingChanges’ also has a profound impact on the model predictions. Multiple booking changes are often represented by blue, suggesting that customers who alter their reservations are less inclined to cancel outright. This could reflect a willingness to adapt plans rather than abandon them.

Also, the variable ‘DepositType_NoDeposit’ stands out on the heatmap, predominantly in red. Not requiring a deposit appears to increase the probability of cancellation. The lack of a financial commitment can make it easier for potential guests to cancel, as they have no monetary loss to deter them. Yet, in a world where bookings and cancellations are more accessible than ever due to the advent of smartphones and smart technology [98], forcing customers to leave a monetary may deter bookings instead of cancellations.

Finally, the choice of bed and breakfast (‘Meal_BB’) has varied impacts on cancellation probabilities. While not as clear-cut as other factors, its presence in the top variables suggests that the type of meal plan could be an indicator of booking stability. This may signal that customers who pay for extra hotel amenities may be less likely to cancel their booking.

The interpretable prediction of cancellations of hotel bookings as a significant contribution to hotel revenue management operations. In particular, several variables such as ‘Country_PRT’, ‘LeadTime’, and ‘DepositType_NoDeposit’ appear to be significant variables pointing to cancellations, suggesting that asking questions such as when, when, and for how much the booking is made are critical for understanding cancellation behaviors of hotel customers. For example, if there is no requirement for a deposit, the risk of cancellation is higher because it becomes unnecessary for any financial involvement to assure that the transaction cannot be altered. On the other hand, the provision of room types ‘AssignedRoomType_A’ and ‘AssignedRoomType_D’ (the most abundant and lowest ADR rooms) appears to be a motivating factor to cancel reservations, meaning that customers that have basic room types may be more likely to cancel than those who book premium rooms.

V. DISCUSSION

The study of hotel booking cancellations has been unpacked through a meticulous examination of predictive models, culminating in the innovation of a meta-model that stands at the confluence of high predictive accuracy and interpretability.

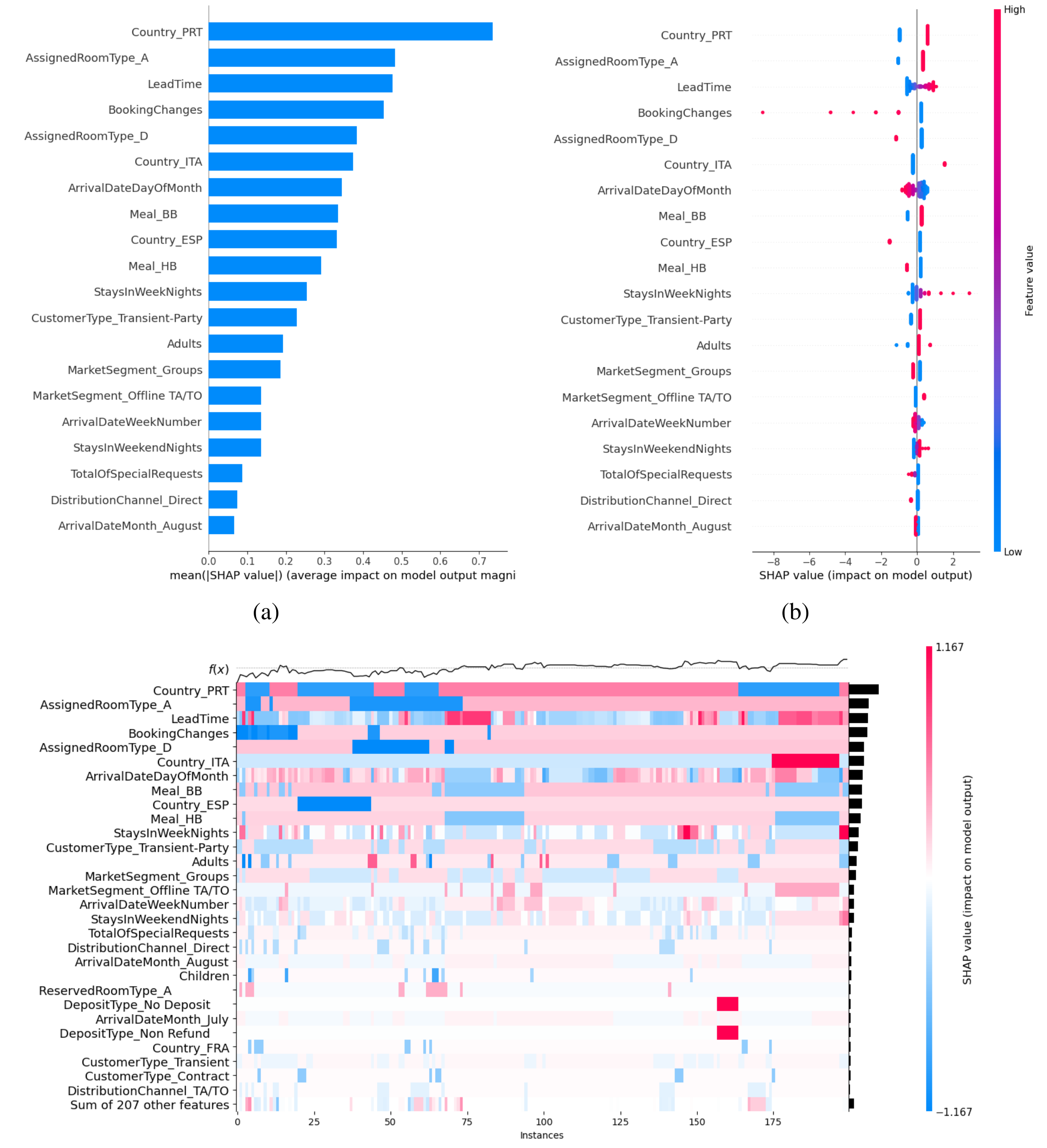


FIGURE 2. Comprehensive SHAP value analysis for the hotel booking cancellation prediction meta-model. Subfigure (a) illustrates the mean SHAP values, demonstrating the average impact of each feature on the model prediction output. Subfigure (b) shows the SHAP value distribution for each feature, highlighting the positive or negative effect on the likelihood of cancellation. Subfigure (c) presents a detailed heatmap of the SHAP values for individual instances of test, indicating the direction and magnitude of each feature influence on specific predictions.

The SHAP value analysis of this study has also shed light on the complex dynamics influencing cancellations, specifically highlighting variables such as country of origin, lead time, and room type as pivotal.

In particular, the SHAP analysis elucidated the significant impact of the country of origin and the lead time on the likelihood of cancelation, aligning with the findings of [20] and [34] that emphasized the importance of the guest origin and planning horizon. The novelty of our approach lies in the composite utilization of these variables within a meta-model framework, enhancing predictive precision beyond the capabilities of individual models explored in previous studies. Moreover, our findings resonate with the emphasis on model interpretability highlighted in [26], where SHAP values served as a cornerstone for transparent decision-making. The actionable insights gleaned from the SHAP

TABLE 4. This table presents a comparative analysis of eight ML models, each subjected to various hyperparameter optimizations, alongside the meta-model developed for forecasting hotel booking cancellations. Noteworthy is the meta-model, marked by an asterisk, which embodies an ensemble methodology that harmonizes the distinct advantages of the individual models, optimized with their respective best hyperparameters.

Model	Hyperparameters	Accuracy	Recall	F1 Score	Precision	Specificity	AUC
LR	'penalty': 'l1', 'C': 0.1	0.8062 ± 0.0031	0.7524 ± 0.0041	0.7637 ± 0.0036	0.7754 ± 0.0057	**0.8445 ± 0.0051**	0.8902 ± 0.0026
	'penalty': 'l1', 'C': 1	**0.8081 ± 0.0025**	**0.7579 ± 0.0054**	**0.7668 ± 0.0033**	**0.7760 ± 0.0043**	0.8439 ± 0.0042	**0.8928 ± 0.0021**
	'penalty': 'l1', 'C': 10	0.8062 ± 0.0037	0.7559 ± 0.0052	0.7645 ± 0.0044	0.7735 ± 0.0064	0.8421 ± 0.0053	0.8923 ± 0.0028
	'penalty': 'l2', 'C': 0.1	0.8056 ± 0.0031	0.7515 ± 0.0058	0.7629 ± 0.0040	0.7747 ± 0.0039	0.8441 ± 0.0038	0.8898 ± 0.0029
	'penalty': 'l2', 'C': 1	0.8071 ± 0.0036	0.7576 ± 0.0060	0.7658 ± 0.0043	0.7743 ± 0.0052	0.8425 ± 0.0047	0.8923 ± 0.0029
	'penalty': 'l2', 'C': 10	0.8069 ± 0.0028	0.7559 ± 0.0063	0.7652 ± 0.0035	0.7748 ± 0.0048	0.8433 ± 0.0044	0.8924 ± 0.0022
SVM	'kernel': 'linear', 'C': 0.1	0.7945 ± 0.0318	0.8978 ± 0.0429	0.7704 ± 0.0299	0.6769 ± 0.0436	0.7311 ± 0.0519	0.8896 ± 0.0239
	'kernel': 'linear', 'C': 10	**0.8450 ± 0.0254**	0.9009 ± 0.0310	**0.8168 ± 0.0277**	**0.7478 ± 0.0343**	**0.8102 ± 0.0325**	**0.9091 ± 0.0106**
	'kernel': 'linear', 'C': 1	0.8140 ± 0.0252	**0.9127 ± 0.0150**	0.7903 ± 0.0228	0.6981 ± 0.0373	0.7530 ± 0.0432	0.8995 ± 0.0161
	'kernel': 'rbf', 'C': 0.1	0.6600 ± 0.0312	0.6248 ± 0.0698	0.5832 ± 0.0448	0.5495 ± 0.0341	0.6823 ± 0.0396	0.7480 ± 0.0346
	'kernel': 'rbf', 'C': 10	0.7110 ± 0.0423	0.8339 ± 0.0759	0.6888 ± 0.0307	0.5923 ± 0.0470	0.6349 ± 0.0881	0.7946 ± 0.0291
	'kernel': 'rbf', 'C': 1	0.7035 ± 0.0348	0.7795 ± 0.0898	0.6673 ± 0.0292	0.5905 ± 0.0412	0.6571 ± 0.0822	0.7855 ± 0.0338
DT	'criterion': 'entropy', 'max_depth': 10	0.8340 ± 0.0044	0.7821 ± 0.0095	0.7969 ± 0.0040	0.8125 ± 0.0133	0.8710 ± 0.0120	0.9131 ± 0.0018
	'criterion': 'entropy', 'max_depth': 5	0.7971 ± 0.0095	0.7272 ± 0.0456	0.7487 ± 0.0072	0.7772 ± 0.0438	0.8468 ± 0.0466	0.8734 ± 0.0033
	'criterion': 'entropy', 'max_depth': None	**0.9114 ± 0.0019**	**0.9205 ± 0.0043**	**0.8964 ± 0.0021**	**0.8735 ± 0.0041**	**0.9049 ± 0.0038**	**0.9158 ± 0.0022**
	'criterion': 'gini', 'max_depth': 10	0.8359 ± 0.0035	0.7915 ± 0.0123	0.8006 ± 0.0035	0.8103 ± 0.0126	0.8676 ± 0.0121	0.9152 ± 0.0024
	'criterion': 'gini', 'max_depth': 5	0.7972 ± 0.0054	0.7590 ± 0.0215	0.7570 ± 0.0043	0.7561 ± 0.0215	0.8244 ± 0.0226	0.8735 ± 0.0033
	'criterion': 'gini', 'max_depth': None	0.9098 ± 0.0027	0.9193 ± 0.0043	0.8946 ± 0.0032	0.8712 ± 0.0044	0.9031 ± 0.0036	0.9143 ± 0.0026
RF	'n_estimators': 50, 'max_depth': 10	0.8407 ± 0.0038	0.7290 ± 0.0094	0.7921 ± 0.0052	0.8673 ± 0.0090	0.9204 ± 0.0067	0.9218 ± 0.0018
	'n_estimators': 50, 'max_depth': 5	0.7993 ± 0.0124	0.5822 ± 0.0528	0.7058 ± 0.0308	0.9036 ± 0.0290	0.9540 ± 0.0197	0.8912 ± 0.0049
	'n_estimators': 50, 'max_depth': None	0.9345 ± 0.0021	0.9211 ± 0.0052	0.9213 ± 0.0026	0.9216 ± 0.0027	0.9441 ± 0.0023	0.9826 ± 0.0007
	'n_estimators': 100, 'max_depth': 10	0.8406 ± 0.0029	0.7242 ± 0.0102	0.7909 ± 0.0053	0.8712 ± 0.0069	0.9236 ± 0.0052	0.9228 ± 0.0020
	'n_estimators': 100, 'max_depth': 5	0.7975 ± 0.0106	0.5737 ± 0.0415	0.7013 ± 0.0265	0.9067 ± 0.0197	**0.9570 ± 0.0123**	0.8930 ± 0.0036
	'n_estimators': 100, 'max_depth': None	0.9349 ± 0.0022	0.9221 ± 0.0032	0.9219 ± 0.0026	0.9216 ± 0.0045	0.9441 ± 0.0034	0.9832 ± 0.0010
	'n_estimators': 200, 'max_depth': 10	0.8415 ± 0.0031	0.7272 ± 0.0063	0.7925 ± 0.0038	0.8709 ± 0.0074	0.9231 ± 0.0054	0.9239 ± 0.0021
	'n_estimators': 200, 'max_depth': 5	0.7995 ± 0.0089	0.5808 ± 0.0383	0.7061 ± 0.0234	0.9044 ± 0.0194	0.9554 ± 0.0127	0.8938 ± 0.0037
	'n_estimators': 200, 'max_depth': None	**0.9354 ± 0.0018**	**0.9231 ± 0.0035**	**0.9224 ± 0.0021**	**0.9218 ± 0.0034**	0.9442 ± 0.0026	**0.9834 ± 0.0008**
GB	'n_estimators': 50, 'learning_rate': 0.1	0.8259 ± 0.0033	0.7691 ± 0.0075	0.7863 ± 0.0044	0.8042 ± 0.0057	0.8665 ± 0.0055	0.9083 ± 0.0018
	'n_estimators': 50, 'learning_rate': 0.01	0.8043 ± 0.0031	0.6637 ± 0.0092	0.7384 ± 0.0049	0.8323 ± 0.0077	0.9046 ± 0.0062	0.8579 ± 0.0082
	'n_estimators': 50, 'learning_rate': 0.001	0.7698 ± 0.0201	0.5392 ± 0.1233	0.6539 ± 0.0550	**0.9015 ± 0.1221**	**0.9344 ± 0.0909**	0.8208 ± 0.0205
	'n_estimators': 100, 'learning_rate': 0.1	0.8338 ± 0.0026	0.7885 ± 0.0046	0.7980 ± 0.0033	0.8077 ± 0.0047	0.8661 ± 0.0039	0.9184 ± 0.0020
	'n_estimators': 100, 'learning_rate': 0.01	0.8041 ± 0.0027	0.6702 ± 0.0090	0.7401 ± 0.0050	0.8264 ± 0.0054	0.8996 ± 0.0044	0.8750 ± 0.0033
	'n_estimators': 100, 'learning_rate': 0.001	0.7859 ± 0.0182	0.5802 ± 0.1166	0.6857 ± 0.0635	0.8861 ± 0.0912	0.9326 ± 0.0558	0.8383 ± 0.0105
	'n_estimators': 200, 'learning_rate': 0.1	**0.8419 ± 0.0025**	**0.8057 ± 0.0050**	**0.8093 ± 0.0030**	0.8129 ± 0.0047	0.8677 ± 0.0044	**0.9262 ± 0.0020**
	'n_estimators': 200, 'learning_rate': 0.01	0.8166 ± 0.0035	0.7579 ± 0.0095	0.7747 ± 0.0047	0.7925 ± 0.0073	0.8584 ± 0.0071	0.8925 ± 0.0029
	'n_estimators': 200, 'learning_rate': 0.001	0.7911 ± 0.0148	0.6241 ± 0.0951	0.7087 ± 0.0517	0.8486 ± 0.0775	0.9099 ± 0.0484	0.8397 ± 0.0066
XGBoost	'n_estimators': 50, 'learning_rate': 0.001	0.8129 ± 0.0059	0.7490 ± 0.0088	0.7692 ± 0.0057	0.7908 ± 0.0142	0.8585 ± 0.0130	0.8883 ± 0.0030
	'n_estimators': 50, 'learning_rate': 0.01	0.8184 ± 0.0038	0.7451 ± 0.0068	0.7736 ± 0.0041	0.8045 ± 0.0099	0.8707 ± 0.0085	0.8952 ± 0.0025
	'n_estimators': 50, 'learning_rate': 0.1	0.8416 ± 0.0028	0.7772 ± 0.0070	0.8034 ± 0.0036	0.8314 ± 0.0050	0.8876 ± 0.0044	0.9254 ± 0.0022
	'n_estimators': 100, 'learning_rate': 0.001	0.8141 ± 0.0062	0.7490 ± 0.0111	0.7703 ± 0.0056	0.7933 ± 0.0156	0.8605 ± 0.0149	0.8895 ± 0.0031
	'n_estimators': 100, 'learning_rate': 0.01	0.8242 ± 0.0035	0.7392 ± 0.0068	0.7778 ± 0.0043	0.8209 ± 0.0080	0.8849 ± 0.0064	0.9008 ± 0.0023
	'n_estimators': 100, 'learning_rate': 0.1	0.8551 ± 0.0032	0.8167 ± 0.0065	0.8244 ± 0.0040	0.8322 ± 0.0054	0.8826 ± 0.0043	0.9370 ± 0.0021
	'n_estimators': 200, 'learning_rate': 0.001	0.8162 ± 0.0063	0.7476 ± 0.0103	0.7720 ± 0.0057	0.7985 ± 0.0164	0.8651 ± 0.0150	0.8915 ± 0.0034
	'n_estimators': 200, 'learning_rate': 0.01	0.8292 ± 0.0028	0.7443 ± 0.0061	0.7839 ± 0.0041	0.8279 ± 0.0044	**0.8897 ± 0.0033**	0.9097 ± 0.0019
	'n_estimators': 200, 'learning_rate': 0.1	**0.8676 ± 0.0032**	**0.8438 ± 0.0050**	**0.8414 ± 0.0035**	**0.8392 ± 0.0043**	0.8847 ± 0.0038	**0.9465 ± 0.0017**
LightGBM	'n_estimators': 50, 'learning_rate': 0.001	0.8284 ± 0.0031	0.7520 ± 0.0082	0.7848 ± 0.0043	0.8208 ± 0.0065	0.8829 ± 0.0061	0.8999 ± 0.0027
	'n_estimators': 50, 'learning_rate': 0.01	0.8296 ± 0.0030	0.7545 ± 0.0070	0.7866 ± 0.0037	0.8217 ± 0.0069	0.8832 ± 0.0057	0.9094 ± 0.0024
	'n_estimators': 50, 'learning_rate': 0.1	0.8552 ± 0.0030	0.8165 ± 0.0043	0.8244 ± 0.0034	0.8325 ± 0.0048	0.8828 ± 0.0040	0.9365 ± 0.0018
	'n_estimators': 100, 'learning_rate': 0.001	0.8288 ± 0.0026	0.7521 ± 0.0086	0.7853 ± 0.0036	0.8216 ± 0.0066	0.8835 ± 0.0061	0.9024 ± 0.0031
	'n_estimators': 100, 'learning_rate': 0.01	0.8334 ± 0.0027	0.7601 ± 0.0064	0.7916 ± 0.0038	0.8259 ± 0.0040	0.8857 ± 0.0033	0.9160 ± 0.0020
	'n_estimators': 100, 'learning_rate': 0.1	0.8656 ± 0.0025	0.8410 ± 0.0057	0.8389 ± 0.0030	0.8369 ± 0.0039	0.8831 ± 0.0035	0.9454 ± 0.0013
	'n_estimators': 200, 'learning_rate': 0.001	0.8290 ± 0.0035	0.7550 ± 0.0093	0.7862 ± 0.0043	0.8201 ± 0.0079	0.8818 ± 0.0073	0.9054 ± 0.0029
	'n_estimators': 200, 'learning_rate': 0.01	0.8410 ± 0.0033	0.7733 ± 0.0060	0.8019 ± 0.0041	0.8328 ± 0.0047	0.8893 ± 0.0036	0.9232 ± 0.0022
	'n_estimators': 200, 'learning_rate': 0.1	**0.8778 ± 0.0022**	**0.8597 ± 0.0042**	**0.8541 ± 0.0026**	**0.8487 ± 0.0041**	**0.8907 ± 0.0034**	**0.9534 ± 0.0015**
MLP	'hidden_layer_sizes': (100,), 'activation': 'logistic'	0.8772 ± 0.0047	0.8892 ± 0.0184	0.8577 ± 0.0039	0.8290 ± 0.0173	0.8686 ± 0.0186	0.9554 ± 0.0015
	'hidden_layer_sizes': (100,), 'activation': 'relu'	0.8653 ± 0.0060	0.8628 ± 0.0349	0.8420 ± 0.0051	0.8246 ± 0.0307	0.8670 ± 0.0323	0.9468 ± 0.0020
	'hidden_layer_sizes': (100,), 'activation': 'tanh'	0.8782 ± 0.0041	0.8767 ± 0.0196	0.8570 ± 0.0037	0.8389 ± 0.0183	0.8793 ± 0.0186	0.9535 ± 0.0015
	'hidden_layer_sizes': (50, 50), 'activation': 'logistic'	0.8669 ± 0.0065	0.8669 ± 0.0316	0.8443 ± 0.0044	0.8249 ± 0.0301	0.8669 ± 0.0316	0.9489 ± 0.0015
	'hidden_layer_sizes': (50, 50), 'activation': 'relu'	0.8695 ± 0.0050	0.8781 ± 0.0193	0.8485 ± 0.0041	0.8216 ± 0.0186	0.8633 ± 0.0199	0.9497 ± 0.0023
	'hidden_layer_sizes': (50, 50), 'activation': 'tanh'	0.8662 ± 0.0045	0.8721 ± 0.0149	0.8444 ± 0.0039	0.8189 ± 0.0154	0.8620 ± 0.0161	0.9478 ± 0.0017
	'hidden_layer_sizes': (100, 50, 25), 'activation': 'logistic'	0.8782 ± 0.0072	0.8749 ± 0.0205	0.8568 ± 0.0051	0.8406 ± 0.0242	0.8806 ± 0.0254	0.9536 ± 0.0015
	'hidden_layer_sizes': (100, 50, 25), 'activation': 'relu'	0.8847 ± 0.0069	0.8928 ± 0.0196	0.8657 ± 0.0060	0.8412 ± 0.0226	0.8789 ± 0.0228	0.9565 ± 0.0019
	'hidden_layer_sizes': (100, 50, 25), 'activation': 'tanh'	**0.8878 ± 0.0054**	**0.8935 ± 0.0115**	**0.8689 ± 0.0057**	**0.8460 ± 0.0138**	**0.8837 ± 0.0130**	**0.9566 ± 0.0021**
(Meta-Model: LR)*	'C': 1, 'solver':liblinear, 'penalty': 'l1'	0.9663 ± 0.0024	0.9593 ± 0.0031	0.9741 ± 0.0019	0.9892 ± 0.0019	0.9797 ± 0.0035	0.9948 ± 0.0025
	'C': 1, 'solver':liblinear, 'penalty': 'l2'	0.9667 ± 0.0022	0.9598 ± 0.0031	0.9744 ± 0.0017	0.9895 ± 0.0015	0.9802 ± 0.0028	0.9948 ± 0.0034
	'C': 1, 'solver':saga, 'penalty': 'l1'	0.9689 ± 0.0010	0.9620 ± 0.0021	**0.9761 ± 0.0008**	**0.9906 ± 0.0011**	**0.9823 ± 0.0021**	0.9951 ± 0.0065
	'C': 1, 'solver':saga, 'penalty': 'l2'	0.9676 ± 0.0021	0.9610 ± 0.0030	0.9751 ± 0.0016	0.9897 ± 0.0015	0.9806 ± 0.0028	0.9950 ± 0.0054
	'C': 10, 'solver':liblinear, 'penalty': 'l1'	0.9686 ± 0.0012	0.9618 ± 0.0016	0.9759 ± 0.0010	0.9904 ± 0.0012	0.9819 ± 0.0023	**0.9952 ± 0.0046**
	'C': 10, 'solver':liblinear, 'penalty': 'l2'	**0.9689 ± 0.0020**	**0.9622 ± 0.0026**	**0.9761 ± 0.0015**	0.9904 ± 0.0008	0.9818 ± 0.0016	0.9951 ± 0.0076
	'C': 10, 'solver':saga, 'penalty': 'l1'	0.9684 ± 0.0021	0.9618 ± 0.0031	0.9758 ± 0.0016	0.9901 ± 0.0014	0.9814 ± 0.0027	0.9949 ± 0.0065
	'C': 10, 'solver':saga, 'penalty': 'l2'	0.9674 ± 0.0012	0.9600 ± 0.0020	0.9749 ± 0.0008	0.9903 ± 0.0017	0.9818 ± 0.0032	0.9947 ± 0.0054

*Note: The meta-model showcased in this table is developed by amalgamating the most efficient models, each fine-tuned with its optimal hyperparameters.

analysis facilitate a nuanced understanding of cancellation drivers, offering hotel managers a detailed roadmap for proactive interventions.

The interpretability of AI models can have a significant impact on the acceptance and use of AI in the hospitality industry, both from an employee and a customer perspective. From the employee perspective, transparency in model decisions can reduce resistance to using advanced systems by shedding light on the reasoning behind decision-making processes. Similarly, consumers may feel a higher sense of fairness and transparency, improving satisfaction and the perception of service received. From an organizational perspective, a company may be able to pivot strategies ethically and effectively, ensuring that business practices are backed by data and align with social and regulatory expectations.

The investigation of hospitality booking cancellations, a subset of the broader field focused on customer churn prediction, has benefited from the integration of advanced ML techniques. The application of AI, as discussed through SHAP value analysis and in our meta-model, exemplifies AI's potential to transform service delivery and RM in the hospitality industry. Our meta-model, which integrates insights from various predictive models including those employed by [20], [24], and [26], addresses the critical need for interpretable and profit-driven analytics in hospitality. Additionally, in practice this machine learning methodology will enable stakeholders to anticipate and adapt to customer behaviors, significantly transforming the landscape of hotel RM and operation.

VI. CONCLUSION, LIMITATIONS, AND FUTURE RESEARCH

This research highlights how effective a state of the art AI model can be in predicting hotel booking cancellations, particularly when coupled with stacking and interpretability techniques. This analysis advances the current state of hotel revenue management literature by aggregating the existing models to increase forecasting accuracy as well as integrating robust explainability metrics to yield actionable insights for hoteliers.

In practical terms, the suggested meta-model can aid operational decision-making as it accurately predicts the cancellation of the bookings and explains the reasons for these cancellations. These insights allow hotel managers to make better decisions regarding overbooking and cancellation policies and booking management. This not only has the potential to affect the hotel's bottom line, but also can help to inform strategy within the organization, particularly when it comes to revenue management, marketing, and loyalty programs.

The model also promotes responsibility in AI based decisions, enabling transparency, fairness, and ethical practices by spelling out the predictions and their rational basis. This information ensures that there is no group that is being systematically targeted by the model, particularly in jurisdictions which lack privacy regulations with respect to consumeer data.

Despite its contributions, this study acknowledges various limitations, including the potential for geographic and market-specific biases in the dataset. Future research directions beckon a broader dataset encompassing varied geographic locations and market segments, further development of model capabilites, and assessment of the willingness of hoteliers to implement AI solutions.

Although various studies have used ML to predict the cancellation of hotel bookings, the models are almost exlusively trained on one source of data [8], [20]. Since the incorporation of other data types to forecasting models has been proven to improve model accuracy in other areas of RM [99], [100], cancellation models may benefit from more diverse datasets. This could be especially true for social media data such as customer reviews or tweets, since cancellation behvaior is affected by social reputation and distribution channel [10].

Another area for future research could be in hotel managers' willingness to implement AI solutions. Although AI has the potential to mitigate many operational challenges, particularly in the realm of revenue management, implementation is not straightforward. Because of the costs associated with servers and platforms for ML modeling and the requirements of technical knowledge from staff that would likely need to get hired, hotel managers may not be as willing or able to implement AI solutions, particularly in smaller operations.

Moreover, the exploration of alternative ensemble techniques and deep learning models may uncover additional insights, offering avenues for further enhancement of predictive performance. Advanced methodologies could be combined with real-time data to produce on-the-spot decisions, allowing assessments to be made as the booking comes in, giving management more time to adapt to bookings with high probabilities of cancellations.

For future work, the insights provided by SHAP on feature importance could be leveraged to guide an iterative model selection process, focusing on refining the experimental framework and further enhancing both model performance and interpretability.

DECLARATION OF COMPETING 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.

REFERENCES

[1] *World Tourism Organization*, World Tourism Organization-UNWTO, Madrid, Spain, 2023.

[2] W. S. Chow, J.-C. Shyu, and K.-C. Wang, "Developing a forecast system for hotel occupancy rate using integrated ARIMA models," *J. Int. Hospitality, Leisure Tourism Manage.*, vol. 1, no. 3, pp. 55–80, Aug. 1998.

[3] F.-L. Chu, "Forecasting tourism demand with ARMA-based methods," *Tourism Manage.*, vol. 30, no. 5, pp. 740–751, Oct. 2009.

[4] D. Frechtling, *Forecasting Tourism Demand*. Evanston, IL, USA: Routledge, 2012.

[5] N. Kourentzes, B. Rostami-Tabar, and D. K. Barrow, "Demand forecasting by temporal aggregation: Using optimal or multiple aggregation levels?" *J. Bus. Res.*, vol. 78, pp. 1–9, Sep. 2017.

[6] S. Li, T. Chen, L. Wang, and C. Ming, "Effective tourist volume forecasting supported by PCA and improved BPNN using Baidu index," *Tourism Manage.*, vol. 68, pp. 116–126, Oct. 2018.

[7] N. Antonio, A. de Almeida, and L. Nunes, "Predicting hotel bookings cancellation with a machine learning classification model," in *Proc. 16th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2017, pp. 1049–1054.

[8] A. J. Sánchez-Medina and E. C-Sánchez, "Using machine learning and big data for efficient forecasting of hotel booking cancellations," *Int. J. Hospitality Manage.*, vol. 89, Aug. 2020, Art. no. 102546.

[9] D. D. Sierag, G. M. Koole, R. D. van der Mei, J. I. van der Rest, and B. Zwart, "Revenue management under customer choice behaviour with cancellations and overbooking," *Eur. J. Oper. Res.*, vol. 246, no. 1, pp. 170–185, Oct. 2015.

[10] N. Antonio, A. de Almeida, and L. Nunes, "Big data in hotel revenue management: Exploring cancellation drivers to gain insights into booking cancellation behavior," *Cornell Hospitality Quart.*, vol. 60, no. 4, pp. 298–319, Nov. 2019.
[11] G. El Hajal and B. Rowson, "The future of hospitality jobs," *Res. Hospitality Manage.*, vol. 10, no. 1, pp. 55–61, Jan. 2020.
[12] M. Li, D. Yin, H. Qiu, and B. Bai, "A systematic review of AI technology-based service encounters: Implications for hospitality and tourism operations," *Int. J. Hospitality Manage.*, vol. 95, May 2021, Art. no. 102930.
[13] P. Limna, "Artificial intelligence (AI) in the hospitality industry: A review article," *Int. J. Comput. Sci. Res.*, vol. 7, pp. 1306–1317, Jan. 2023.
[14] D. Gursoy and R. Cai, "Artificial intelligence: An overview of research trends and future directions," *Int. J. Contemp. Hospitality Manage.*, vol. 37, no. 1, pp. 1–17, Jan. 2025.
[15] H. Lv, S. Shi, and D. Gursoy, "A look back and a leap forward: A review and synthesis of big data and artificial intelligence literature in hospitality and tourism," *J. Hospitality Marketing Manage.*, vol. 31, no. 2, pp. 145–175, Feb. 2022.
[16] N. N. Y. Vo, S. Liu, X. Li, and G. Xu, "Leveraging unstructured call log data for customer churn prediction," *Knowl.-Based Syst.*, vol. 212, Jan. 2021, Art. no. 106586.
[17] K. A. Amuda and A. B. Adeyemo, "Customers churn prediction in financial institution using artificial neural network," 2019, *arXiv:1912.11346*.
[18] A. Kasem Ahmad, A. Jafar, and K. Aljoumaa, "Customer churn prediction in telecom using machine learning and social network analysis in big data platform," 2019, *arXiv:1904.00690*.
[19] A. Dhini and M. Fauzan, "Predicting customer churn using ensemble learning: Case study of a fixed broadband company," *Int. J. Technol.*, vol. 12, no. 5, p. 1030, Dec. 2021.
[20] N. Antonio, A. D. Almeida, and L. Nunes, "Predicting hotel booking cancellations to decrease uncertainty and increase revenue," *Tourism Manage. Stud.*, vol. 13, no. 2, pp. 25–39, Apr. 2017.
[21] C. Lemke, S. Riedel, and B. Gabrys, "Evolving forecast combination structures for airline revenue management," *J. Revenue Pricing Manage.*, vol. 12, no. 3, pp. 221–234, May 2013.
[22] S. S. Azadeh, R. Labib, and G. Savard, "Railway demand forecasting in revenue management using neural networks," *Int. J. Revenue Manage.*, vol. 7, no. 1, p. 18, 2013.
[23] D. Romero Morales and J. Wang, "Forecasting cancellation rates for services booking revenue management using data mining," *Eur. J. Oper. Res.*, vol. 202, no. 2, pp. 554–562, Apr. 2010.
[24] N. António, A. de Almeida, and L. M. M. Nunes, "Using data science to predict hotel booking cancellations," in *Handbook of Research on Holistic Optimization Techniques in the Hospitality, Tourism, and Travel Industry*. Hershey, PA, USA: IGI Global, 2017, pp. 141–167.
[25] A. A. Nababan, M. Jannah, and A. H. Nababan, "Prediction of hotel booking cancellation using K-nearest neighbors (K-Nn) algorithm and synthetic minority over-sampling technique (Smote)," *INFOKUM*, vol. 10, no. 3, pp. 50–56, 2022.
[26] Z. Liu, P. Jiang, J. Wang, Z. Du, X. Niu, and L. Zhang, "Hospitality order cancellation prediction from a profit-driven perspective," *Int. J. Contemp. Hospitality Manage.*, vol. 35, no. 6, pp. 2084–2112, May 2023.
[27] Y. Y. Febrian, D. R. Wijaya, and E. Ervina, "Hotel reservation cancellation prediction using boosting model," in *Proc. 2nd Int. Conf. Softw. Eng. Inf. Technol. (ICoSEIT)*, Feb. 2024, pp. 138–143.
[28] Z. A. Andriawan, S. R. Purnama, A. S. Darmawan, Ricko, A. Wibowo, A. Sugiharto, and F. Wijayanto, "Prediction of hotel booking cancellation using CRISP-DM," in *Proc. 4th Int. Conf. Informat. Comput. Sci. (ICICoS)*, Nov. 2020, pp. 1–6.
[29] I. Gómez-Talal, L. González-Serrano, J. L. Rojo-Álvarez, and P. Talón-Ballestero, "Avoiding food waste from restaurant tickets: A big data management tool," *J. Hospitality Tourism Technol.*, vol. 15, no. 2, pp. 232–253, Mar. 2024.
[30] G. Sekhon and S. Ahuja, "Review machine learning models for managing hotel cancellations in the tourism industry," in *Proc. 3rd Int. Conf. Intell. Technol. (CONIT)*, Jun. 2023, pp. 1–6.
[31] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 785–794.
[32] M. Yoo, A. K. Singh, and N. Loewy, "Predicting hotel booking cancelation with machine learning techniques," *J. Hospitality Tourism Technol.*, vol. 15, no. 1, pp. 54–69, Jan. 2024.
[33] C.-C. Chen, Z. Schwartz, and P. Vargas, "The search for the best deal: How hotel cancellation policies affect the search and booking decisions of deal-seeking customers," *Int. J. Hospitality Manage.*, vol. 30, no. 1, pp. 129–135, Mar. 2011.
[34] H. Nanang, "Exploratory data analysis & booking cancelation prediction on hotel booking demands datasets," *J. Appl. Data Sci.*, vol. 2, no. 1, pp. 40–56, Jan. 2021.
[35] C. T. Voulgaris, "Crystal balls and black boxes: What makes a good forecast?" *J. Planning Literature*, vol. 34, no. 3, pp. 286–299, Aug. 2019.
[36] M. Falk and M. Vieru, "Modelling the cancellation behaviour of hotel guests," *Int. J. Contemp. Hospitality Manage.*, vol. 30, no. 10, pp. 3100–3116, Oct. 2018.
[37] A. Riasi, Z. Schwartz, and C.-C. Chen, "A proposition-based theorizing approach to hotel cancellation practices research," *Int. J. Contemp. Hospitality Manage.*, vol. 30, no. 11, pp. 3211–3228, Nov. 2018.
[38] C.-C. Chen and K. L. Xie, "Differentiation of cancellation policies in the U.S. hotel industry," *Int. J. Hospitality Manage.*, vol. 34, pp. 66–72, Sep. 2013.
[39] L. Masiero, J. L. Nicolau, and R. Law, "A demand-driven analysis of tourist accommodation price: A quantile regression of room bookings," *Int. J. Hospitality Manage.*, vol. 50, pp. 1–8, Sep. 2015.
[40] K. A. McGuire, *Hotel Pricing in a Social World: Driving Value in the Digital Economy*. Hoboken, NJ, USA: Wiley, 2015.
[41] N. K. Hikmawati, Y. Ramdhani, and W. Wartika, "Exploring ADR trends: A data mining approach to hotel room pricing, cancellations, and EDA," *J. Appl. Data Sci.*, vol. 5, no. 1, pp. 189–202, 2024.
[42] A. Garrido-Moreno, V. J. García-Morales, and R. Martín-Rojas, "Going beyond the curve: Strategic measures to recover hotel activity in times of COVID-19," *Int. J. Hospitality Manage.*, vol. 96, Jul. 2021, Art. no. 102928.
[43] M. A.-U.-D. Malik, "COVID-19: Global macroeconomics," *J. Home Page*, vol. 3, no. 1, pp. 119–134, Dec. 2021.
[44] A. Herrera, Á. Arroyo, A. Jiménez, and Á. Herrero, "Forecasting hotel cancellations through machine learning," *Expert Syst.*, vol. 41, no. 9, p. 13608, Sep. 2024.
[45] M. Song, W. Wu, and Y. Duan, "The sequelae of hotel pre-sale: The influence of electronic word-of-mouth dispersion on booking cancellation behavior," in *Proc. IEEE/ACIS 19th Int. Conf. Comput. Inf. Sci. (ICIS)*, Jun. 2021, pp. 3–9.
[46] W. Verbeke, K. Dejaeger, D. Martens, J. Hur, and B. Baesens, "New insights into churn prediction in the telecommunication sector: A profit driven data mining approach," *Eur. J. Oper. Res.*, vol. 218, no. 1, pp. 211–229, Apr. 2012.
[47] D. J. Hand, "Measuring classifier performance: A coherent alternative to the area under the ROC curve," *Mach. Learn.*, vol. 77, no. 1, pp. 103–123, Oct. 2009.
[48] D. H. Wolpert, "Stacked generalization," *Neural Netw.*, vol. 5, no. 2, pp. 241–259, Jan. 1992.
[49] J. Sill, G. Takacs, L. Mackey, and D. Lin, "Feature-weighted linear stacking," 2009, *arXiv:0911.0460*.
[50] U. R. Bhosale, M. S. Sutaone, and M. Narkhede, "Empirical study of stacked model for predicting various data," in *Proc. Int. Conf. Signal Inf. Process. (IConSIP)*, Aug. 2022, pp. 1–4.
[51] F. Gunes, R. Wolfinger, and P.-Y. Tan, "Stacked ensemble models for improved prediction accuracy," SAS Inst. Inc., Cary, NC, USA, Tech. Rep.,2017.
[52] Md. S. Satu, K. Ahammed, and M. Z. Abedin, "Performance analysis of machine learning techniques to predict hotel booking cancellations in hospitality industry," in *Proc. 23rd Int. Conf. Comput. Inf. Technol. (ICCIT)*, Dec. 2020, pp. 1–6.
[53] S. Chen, E. W. T. Ngai, Y. Ku, Z. Xu, X. Gou, and C. Zhang, "Prediction of hotel booking cancellations: Integration of machine learning and probability model based on interpretable feature interaction," *Decis. Support Syst.*, vol. 170, Jul. 2023, Art. no. 113959.
[54] P. K. Shukla and S. P. Tripathi, "A survey on interpretability-accuracy (I–A) trade-off in evolutionary fuzzy systems," in *Proc. 5th Int. Conf. Genetic Evol. Comput.*, Aug. 2011, pp. 97–101.
[55] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, "A survey of methods for explaining black box models," *ACM Comput. Surveys*, vol. 51, no. 5, pp. 1–42, Sep. 2019.
[56] J. Torresen, "A review of future and ethical perspectives of robotics and AI," *Frontiers Robot. AI*, vol. 4, p. 75, Jan. 2018.

[57] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," 2017, *arXiv:1702.08608*.

[58] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018.

[59] T. Miller, "Explanation in artificial intelligence: Insights from the social sciences," *Artif. Intell.*, vol. 267, pp. 1–38, Feb. 2019.

[60] A. Rai, "Explainable AI: From black box to glass box," *J. Acad. Marketing Sci.*, vol. 48, no. 1, pp. 137–141, Jan. 2020.

[61] A. Jobin, M. Ienca, and E. Vayena, "The global landscape of AI ethics guidelines," *Nature Mach. Intell.*, vol. 1, no. 9, pp. 389–399, Sep. 2019.

[62] S. Lundberg and S. Lee, "A unified approach to interpreting model predictions," in *Proc. Adv. Neural Inf. Process. Syst.*, Jan. 2017, pp. 4765–4774.

[63] B. Mittelstadt, C. Russell, and S. Wachter, "Explaining explanations in AI," in *Proc. Conf. Fairness, Accountability, Transparency*, Jan. 2019, pp. 279–288.

[64] R. Cai, L. N. Cain, and H. Jeon, "Customers' perceptions of hotel AI-enabled voice assistants: Does brand matter?" *Int. J. Contemp. Hospitality Manage.*, vol. 34, no. 8, pp. 2807–2831, Jul. 2022.

[65] P. Rodríguez, M. A. Bautista, J. Gonzalez, and S. Escalera, "Beyond one-hot encoding: Lower dimensional target embedding," *Image Vis. Comput.*, vol. 75, pp. 21–31, Jul. 2018.

[66] B. Bengfort, R. Bilbro, and T. Ojeda, *Applied Text Analysis With Python: Enabling Language-Aware Data Products With Machine Learning*. Sebastopol, CA, USA: O'Reilly Media, Inc., 2018.

[67] J. Li, Y. Si, T. Xu, and S. Jiang, "Deep convolutional neural network based ECG classification system using information fusion and one-hot encoding techniques," *Math. Problems Eng.*, vol. 2018, pp. 1–10, Dec. 2018.

[68] J. Brownlee, *Introduction To Time Series Forecasting With Python: How To Prepare Data and Develop Models To Predict the Future*. Machine Learning Mastery, 2017.

[69] C. Seger, "An investigation of categorical variable encoding techniques in machine learning: Binary versus one-hot and feature hashing," Tech. Rep., 2018.

[70] L. Breiman, *Classification and Regression Trees*. Evanston, IL, USA: Routledge, 2017.

[71] A. Liaw and M. Wiener, "Classification and regression by randomforest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.

[72] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," *Frontiers neurorobotics*, vol. 7, p. 21, Dec. 2013.

[73] M. Dia, N. Macris, F. Krzakala, T. Lesieur, and L. Zdeborová, "Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 1–9.

[74] Y. Shi, G. Ke, Z. Chen, S. Zheng, and T.-Y. Liu, "Quantized training of gradient boosting decision trees," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 35, 2022, pp. 18822–18833.

[75] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–9.

[76] M. Barton and B. Lennox, "Model stacking to improve prediction and variable importance robustness for soft sensor development," *Digit. Chem. Eng.*, vol. 3, Jun. 2022, Art. no. 100034.

[77] A. Chatzimparmpas, R. M. Martins, K. Kucher, and A. Kerren, "Stack-GenVis: Alignment of data, algorithms, and models for stacking ensemble learning using performance metrics," *IEEE Trans. Vis. Comput. Graphics*, vol. 27, no. 2, pp. 1547–1557, 2020.

[78] M. Claesen and B. De Moor, "Hyperparameter search in machine learning," 2015, *arXiv:1502.02127*.

[79] B. Sahu, S. Agrawal, H. Dey, and C. Raj, "Performance analysis of state-of-the-art classifiers and stack ensemble model for liver disease diagnosis," in *Biologically Inspired Techniques in Many Criteria Decision Making*. Cham, Switzerland: Springer, 2022, pp. 95–105.

[80] J. C. Stoltzfus, "Logistic regression: A brief primer," *Academic Emergency Med.*, vol. 18, no. 10, pp. 1099–1104, Oct. 2011.

[81] C. Schaffer, "Cross-validation," in *Selecting Models From Data: Artificial Intelligence and Statistics IV*. Cham, Switzerland: Springer, 1994, pp. 51–59.

[82] T.-T. Wong and N.-Y. Yang, "Dependency analysis of accuracy estimates in k-Fold cross validation," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 11, pp. 2417–2427, Nov. 2017.

[83] X. Zeng and T. R. Martinez, "Distribution-balanced stratified cross-validation for accuracy estimation," *J. Experim. Theor. Artif. Intell.*, vol. 12, no. 1, pp. 1–12, Jan. 2000.

[84] T.-T. Wong, "Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation," *Pattern Recognit.*, vol. 48, no. 9, pp. 2839–2846, Sep. 2015.

[85] P. Swarnalatha and B. K. Tripathy, "Evaluation of classifier models using stratified tenfold cross validation techniques," in *Proc. Int. Conf. Comput. Commun. Syst.*, Jan. 2012, pp. 680–690.

[86] C. M. Bishop and N. M. Nasrabad, *Pattern Recognition and Machine Learning*, vol. 4. Cham, Switzerland: Springer, 2006.

[87] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006.

[88] D. M. W. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation," 2020, *arXiv:2010.16061*.

[89] L. S. Shapley, "A value for n-person games," Princeton Univ. Press Princeton, Princeton, NJ, USA, Tech. Rep., 1953.

[90] A. E. Roth, *The Shapley Value: Essays in Honor of Lloyd S. Shapley*. Cambridge, U.K.: Cambridge Univ. Press, 1988.

[91] L. Merrick and A. Taly, "The explanation game: Explaining machine learning models using Shapley values," in *Proc. Int. Cross-Domain Conf. Mach. Learn. Knowl. Extraction*, Dublin, Ireland. Cham, Switzerland: Springer, Jan. 2020, pp. 17–38.

[92] H. Duan, "Computational methods for estimating global sensitivity indices Shapley values," Ph.D thesis, Florida State Univ., Tallahassee, FL, USA, 2023.

[93] I. E. Kumar, S. Venkatasubramanian, C. Scheidegger, and S. A. Friedler, "Problems with Shapley-value-based explanations as feature importance measures," in *Proc. Int. Conf. Mach. Learn.*, Jan. 2020, pp. 5491–5500.

[94] G. Laberge and Y. Pequignot, "Understanding interventional TreeSHAP: How and why it works," 2022, *arXiv:2209.15123*.

[95] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee, "From local explanations to global understanding with explainable AI for trees," *Nature Mach. Intell.*, vol. 2, no. 1, pp. 56–67, Jan. 2020.

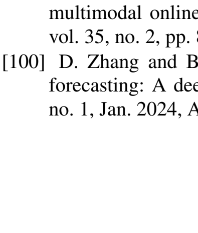
[96] K. C. Lichtendahl and R. L. Winkler, "Why do some combinations perform better than others?" *Int. J. Forecasting*, vol. 36, no. 1, pp. 142–149, Jan. 2020.

[97] O. Sagi and L. Rokach, "Ensemble learning: A survey," *WIREs Data Mining Knowl. Discovery*, vol. 8, no. 4, p. 1249, Jul. 2018.

[98] H. Lee, S.-B. Yang, and N. Chung, "Out of sight," *J. Air Transp. Manage.*, vol. 90, Jan. 2021, Art. no. 101942.

[99] G. Chen, L. Huang, S. Xiao, C. Zhang, and H. Zhao, "Attending to customer attention: A novel deep learning method for leveraging multimodal online reviews to enhance sales prediction," *Inf. Syst. Res.*, vol. 35, no. 2, pp. 829–849, Jun. 2024.

[100] D. Zhang and B. Niu, "Leveraging online reviews for hotel demand forecasting: A deep learning approach," *Inf. Process. Manage.*, vol. 61, no. 1, Jan. 2024, Art. no. 103527.



ISMAEL GÓMEZ-TALAL (Member, IEEE) received the B.Sc. degree in telecommunication technologies engineering and the M.Sc. degree in information systems engineering from the Universidad Rey Juan Carlos (URJC), in 2020 and 2021, respectively, and the M.Sc. degree in Internet of Things from the Universidad Carlos III de Madrid (UC3M), in 2022. He is currently pursuing the degree in telematics engineering with URJC. His research interests include methods based on interpretable machine learning (IML) and the application of machine learning (ML) using real data from the health and hospitality sectors.

MANA AZIZSOLTANI received the bachelor's degree in mathematics from the University of Nevada, Las Vegas (UNLV) and the master's degree in statistics from North Carolina State University (NCSU). He is currently pursuing the Ph.D. degree in hospitality administration. He is also studying applications of machine learning in the hotel and casino industry for the Ph.D. degree. He is also a Research Scientist with the International Gaming Institute, UNLV, where he conducts research on the applications of artificial intelligence for responsible gambling and casino marketing. He is also the Head of Data Science with the Differential Laboratories, where he assists casino operators and suppliers develop and implement data science solutions.

PILAR TALÓN-BALLESTERO is currently a Ph.D. Holder in advanced marketing, an Associate Professor with the Business Economics Department, and the Director of University Master in Revenue Management, Rey Juan Carlos University. She is also the Director and a Researcher in numerous studies and consultancy projects developed with private and public entities: Spanish Confederation of Hotels and Touristic Accommodations, Spanish Hotel Technological Institute, Iberostar, NH, and Palladium Hotels Groups. She has published numerous publications (articles and books), including some impact ones about hotel sector. She was the Hotel and Travel Agency Manager. Her areas of specialization are revenue management, distribution, big data, and gender.

ASHOK SINGH is currently an Applied Statistics and Gaming Odds Expert who teaches UNLV courses on the math of casino games, advanced statistics, slot analytics, design, mathematical models for business, and operations research. He additionally leads a graduate-level course on data mining and machine learning in the software environment R, a programming language for statistical computing, and graphics. He joined UNLV, in 1991, as a Faculty Member with the Department of Mathematical Sciences, where he developed and taught undergraduate and graduate courses in statistics. He transferred to the College of Hospitality, Gaming Faculty, in 2006, and was named as the Chair of the Department of Resorts, Gaming and Golf Management, in 2021. In addition to teaching, he is also a Researcher whose interests span a wide range of topics, including medical research, public health, and civil and environmental engineering, including transportation systems to support research projects sponsored by local, state, and federal agencies and the private sector. His most recent projects have examined the impacts of a smoking ban on casino gaming volume, the relationship between the house edge and play time, and casino loyalty programs.