The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/0959-6119.htm

Received 4 May 2022
Revised 23 August 2022
14 November 2022
10 January 2023
Accepted 31 January 2023

An ensemble machine learning framework for Airbnb rental price modeling without using amenity-driven features

Indranil Ghosh
*IT and Analytics Area,
Institute of Management Technology Hyderabad, Hyderabad, India*
Rabin K. Jana
*Operations and Quantitative Methods Area,
Indian Institute of Management Raipur, Raipur, India, and*
Mohammad Zoynul Abedin
*Department of Finance, Performance and Marketing,
Teesside University International Business School, Middlesbrough, UK*

Abstract

Purpose – The prediction of Airbnb listing prices predominantly uses a set of amenity-driven features. Choosing an appropriate set of features from thousands of available amenity-driven features makes the prediction task difficult. This paper aims to propose a scalable, robust framework to predict listing prices of Airbnb units without using amenity-driven features.

Design/methodology/approach – The authors propose an artificial intelligence (AI)-based framework to predict Airbnb listing prices. The authors consider 75 thousand Airbnb listings from the five US cities with more than 1.9 million observations. The proposed framework integrates (i) feature screening, (ii) stacking that combines gradient boosting, bagging, random forest, (iii) particle swarm optimization and (iv) explainable AI to accomplish the research objective.

Findings – The key findings have three aspects – prediction accuracy, homogeneity and identification of best and least predictable cities. The proposed framework yields predictions of supreme precision. The predictability of listing prices varies significantly across cities. The listing prices are the best predictable for Boston and the least predictable for Chicago.

Practical implications – The framework and findings of the research can be leveraged by the hosts to determine rental prices and augment the service offerings by emphasizing key features, respectively.

Originality/value – Although individual components are known, the way they have been integrated into the proposed framework to derive a high-quality forecast of Airbnb listing prices is unique. It is scalable. The Airbnb listing price modeling literature rarely witnesses such a framework.

Keywords Airbnb, Listing price, Ensemble machine learning, Stacking, Explainable AI

Paper type Research paper

Introduction

Airbnb has emerged as a significant player in the accommodation sharing-economy market (Qiu *et al.*, 2022). Over ten million hosts provide rental services across 30,000 cities

The authors are grateful to the anonymous reviewers, the Editor-in-Chief and the Associate Editor for their careful evaluation to improve the quality and presentation of the paper.

0959-6119
DOI 10.1108/IJCHM-05-2022-0562

worldwide via Airbnb. The peer-to-peer (P2P) accommodation service is increasing rapidly (Farmaki *et al.*, 2021; Zhu *et al.*, 2021). Among various factors, societal and economic considerations are the primary drivers of P2P accommodation services (Tussyadiah, 2015; Zhu *et al.*, 2019; Liu *et al.*, 2022). Visitors seek secure and cost-effective accommodations (Guttentag, 2015). This rapid boom of P2P accommodation services motivates researchers to investigate various aspects of Airbnb listing price modeling (So *et al.*, 2022). Listing price is of paramount relevance for the sustainability of P2P accommodation in the long run (Han and Bai, 2022). The generic offering-related features explain the variation in listing price (Lawani *et al.*, 2019; Zhu *et al.*, 2021). These studies focus on testing research hypotheses for evaluating the influence of respective features without focusing on the predictive modeling of the listing prices. On the other hand, authors have predicted Airbnb listing prices using amenities-driven features (Kalehbasti *et al.*, 2021; Islam *et al.*, 2022). Thus, no focus has been given to Airbnb listing price modeling without using amenity-driven features. Also, selecting the most appropriate feature from thousands of available amenity-driven features makes the Airbnb listing price modeling task difficult. We bridge this gap by proposing an ensemble machine learning (ML) framework for Airbnb listing price modeling without using amenity-driven features.

The major contributions of this research are three folds. First, the empirical relationship between the explanatory constructs and the listing prices is often interlocked through a nonlinear bond that complicates discovering the patterns. We develop a robust predictive structure using a set of nonamenity-driven features to forecast listing prices of nearly 75 thousand Airbnb units located in five US cities. Second, we contribute methodologically to the Airbnb listing price forecasting literature by proposing a novel scalable framework consisting of feature selection, nature-inspired optimization, ensemble ML and Explainable artificial intelligence (AI). Third, despite the existence of past cognate literature on the evaluation of the effect of various related features on the overall experience and listing prices, most of these works are either confined to survey-based research hypothesis testing or deploying regression-based structures to draw inferences. The entire gamut of the said literature stands on a preconceived notion of the linear and parametric form of relationship. We use Explainable AI to interpret the predictive ensemble modeling for forecasting listing prices to ascertain the nature of the impact of chosen features in a nonparametric setup. Findings of the Explainable AI would be of practical implications in the context of regulation and augmentation of service quality of the Airbnb platform.

In this research, we select 25 explanatory features of numeric and factor variables pertinent to essential aspects of accommodation, booking and host characteristics. We do not incorporate appliance-related amenities, as it will enormously increase the number of independent variables and computational complexity. The Boruta feature selection algorithm evaluates the explanatory capability of features for filtering out the irrelevant ones. Three ensemble predictive ML algorithms, namely, random forest (RF), bagging and gradient boosting (GB), are used for forecasting rental rates. The stacking algorithm pools these predicted listing prices to derive the final prediction. This algorithm learns through extreme gradient boosting (XGBoost). Thus, we obtain the final forecast of Airbnb units' listing prices using an ensemble of ensemble framework. The performances of all these frameworks are sensitive to hyper-parameter tuning. We use the particle swarm optimization (PSO) algorithm for this purpose. Finally, Explainable AI is invoked to decode the ensemble forecasting framework to convert the computational model into practical relevance. Demystifying the influence pattern of the offering related independent features through Explainable AI is highly useful to fathom the customers' voice in the tourism sharing economy platform. Therefore, the findings of the study reveal critical inferences pertinent to the dynamics of Airbnb rental prices and the trend of consumer preference in the sharing economy platform. The hosts can leverage the research findings to set pricing for maximizing revenue by appropriately emphasizing the most significant features. The integration of Explainable AI pimps into AI-based methodology serves deeper insights, which adds novelty to the current work compared to the existing AI-based literature pertinent to tourism and hospitality (Doborjeh *et al.*, 2022; Mariani and Baggio, 2022). Hence, the present research contributes to the methodological front and transpires to be practically relevant by serving insights for pricing strategy.

Literature review

Throughout the literature, different strands of research on listing prices and overall satisfaction with staying at Airbnb units have seen serious traction among researchers (Dogru *et al.*, 2021). Most of the AI-driven previous research assesses the influence of determinants on listing prices, revenues and customer satisfaction (Gyodi and Nawaro, 2021). Sainaghi (2021) conducted an exhaustive literature survey to document the set of explanatory constructs of the price and revenue of P2P accommodation platforms. He observed that the listing variables have primarily been identified as the major determinants. The said findings rationalize the endeavor of the present work to build a practically implementable framework for predicting Airbnb listing prices. Wang and Nicolau (2017) deployed orthodox and quantile regression models to establish the prevailing relationship between listing prices and their determinants across 33 cities. The results indicated that amenities, online ratings, rental rules, property characteristics and host attributes largely influence the listing prices.

Lawani *et al.* (2019) used the hedonic spatial autoregressive and sentiment scoring models to decode the impacts of reviews, various amenities and neighborhood-related features on Airbnb listing prices. They show that the review score with room amenities and neighborhood features influence listing prices. Also, policies undertaken to improve the quality of a host resulted in a spillover effect on the competitor hosts' listing prices. Hu *et al.* (2019) identify contributing factors of housing listing prices using ML and social media data. Lagonigro *et al.* (2020) identified the key factors responsible for the variation of Airbnb listing prices in Barcelona, Spain, using geographically weighted regression. Chica-Olmo *et al.* (2020) inspected the predictive dependence induced by features linked to amenities, the behavior of the host and locational aspects of Airbnb units' rental prices in the Malaga region of Spain. Sainaghi *et al.* (2021) leveraged hedonic modeling to identify the primary factors of Airbnb rental prices in Milan, Italy. It was revealed that the constructs pertinent to location, size, type of listing and seasonality significantly explain the variability of rental prices in heterogeneous categories of property.

ML-driven modeling of Airbnb rental prices has garnered traction in the academic fraternity. Song *et al.* (2022) emphasized the need to incorporate ML modeling to explore the deeper dynamics of Airbnb offerings. Liu (2021) applied a series of regularized and ensemble regression methodologies to predict Airbnb rental prices. The XGBoost method outperformed the competing models in terms of prediction accuracy. Yang (2021) used XGBoost and the artificial neural network for a successful predictive analysis of Airbnb rental prices. The findings identified key amenities that can be leveraged to increase rental prices. Islam *et al.* (2022) combined sentiment analysis and ensemble ML to predict Airbnb rental prices by incorporating textual information, spatial features and amenities as explanatory variables. Thakur *et al.* (2022) performed predictive modeling of Airbnb units in Rio de Janeiro using a deep neural network to identify the influential features. Combinations of hedonic modeling and quantitative techniques for discovering the rental price determinants of Airbnb units have been documented (Sainaghi *et al.*, 2021; Lee *et al.*, 2022; Wang and Rasouli, 2022). Text mining approaches have also been used to comprehend related issues driving rental prices (Yu *et al.*, 2021; Zhang *et al.*, 2022).

Analysis of the literature indicates a strong trend among the researchers to identify the critical factors affecting listing prices. Most existing research uses primary data retrieved by administering a survey or secondary data on the Airbnb listing sites. All these works generate insights to enable the hosts to undertake several measures for improving their service and offerings. It provides regulatory frameworks as well for setting up appropriate listing rates. There is an apparent shortage of work to develop predictive structures to estimate precise forecasts of listing prices in a scalable manner. It is important to note that plenteous attributes are interlinked with listing prices. Considering all possible features for predictive modeling is practically infeasible. On the other hand, hedonic model-driven attribute assessment to ascertain their role suffers from assuming the linear form of relationship. Thus, developing a scalable predictive model and following it up through a systematic model explanation to infer the nature and direction of influence can effectively void the prevailing research gap and serve critical insights.

From the methodological front, the usage of ensemble ML for estimating rental prices has been reported in the literature. Nevertheless, the existing literature does not fully use the efficacy and potential ensemble ML in constructing a practically implementable framework for modeling dynamics of the listing price in different cities with a systematic evaluation of the offering-related features. The current research resorts to leveraging them with nature-inspired optimization algorithms to deduce forecasts of supreme precision with affordable computation expenses. Despite the previous attempts to explore the predictability of rental prices using ML algorithms, the said studies offer little contribution to model interpretation. Ensemble ML techniques provide high accuracy at the expense of explainability. The current work bridges the gap by applying Explainable AI on top of the proposed predictive framework to infer significant insights on the influence of explanatory features.

Data description

We consider approximately 75 thousand Airbnb listings across five US cities – Austin, Boston, Chicago, Los Angeles and Nashville consisting of 1.919164 million observations to accomplish our research endeavors. To avoid the ongoing COVID-19 pandemic, we choose data from November 2019 on listing prices and 25 distinct features – Superhost, Host_Verified, Host_listings, Accommodates, Room_Type, Property_Type, Bedrooms, Bathrooms, Beds, Extra_People charge, Maximum_Night, Minimum_Night, Review_Scores, Review_12 (total reviews in previous 12-months), Availability_30, Availability_60, Availability_365 (rooms available for the coming 30, 60 and 365 days), Reviews, instant booking status (Booking), cancellation policy (Cancellation), requirement of guest picture (Guest_Picture), total host listings (Total_Listings) and reviews per month (Review_M) have been compiled from the data repository portal of Airbnb, http://insideairbnb.com/get-the-data.html. Many of these variables are a factor in nature that requires the creation of dummy variables to build the predictive framework. Information on a plethora of amenities is also available. However, including such features in the predictive modeling framework is impossible as the number of such features will cross the number of data points for a particular city.

We have computed the various measures of central tendency and dispersion of Airbnb rental prices in select locations. Out of these measures, the standard deviation for Boston is the least. Hence, the listing price of Airbnb units across Boston is comparatively more uniform than their counterparts. We now introspect whether the rental price distribution is uniform across the locations. To accomplish the task, we used the Kolmogorov–Smirnov test and Mann–Whitney tests. The results show that all the test statistics values are highly significant for all five cities. Thus, Airbnb units' prices are nonuniform. Hence, the importance of various features will vary location-wise. Therefore, developing a scalable predictive structure to capture the rental price variations across cities will be challenging. It also rationalizes deploying a dedicated feature selection algorithm to critically identify the significant features, as applying all features in a uniform setup can inhibit prediction accuracy. Then, we explore the nature of the association between the considered variables, which indicates the absenteeism of orthodox correlation. The presence of multicollinearity is minimal too. Therefore, deploying methods capable of mining complex nonlinear associations through the Boruta feature selection algorithm and subsequent ensemble modeling for predictive analysis is justified. We use Kruskal–Wallis's one-way analysis of variance (ANOVA) to check the factor variables' association structure. Table 1 narrates the test results.

The test statistics values are highly significant in most cases barring a few instances. Hence, we cannot rule out the impact of these features in driving the pricing strategies. We consider all numerical and factor features for evaluation through the dedicated nonlinear feature selection procedures of Boruta. This research applies a one-hot encoding scheme to transform the factor features into numerical ones, where $(n-1)$ dummy variables are created from a factor containing n categories.

Methodology

We briefly present different components of the methodology adopted in the proposed framework.

Boruta feature selection algorithm

Boruta is based on an ensemble learning paradigm (Kursa and Rudnicki, 2010). It follows the operational procedure of RF with updated steps for ranking the importance of explanatory variables. Boruta incorporates an amplified randomness level in the existing system to select features with significant explanatory capabilities. It is helpful in resolving supervised feature screening tasks (Jana *et al.*, 2020). The algorithm helps to rank the 25 considered nonamenity features related to the Airbnb units of five chosen US cities.

Particle swarm optimization

The PSO is a stochastic search method (Kennedy and Eberhart, 1995). The method works efficiently for complex optimization problems (Jana *et al.*, 2021; Pradhan *et al.*, 2021). We use

Table 1. Outcome of the Kruskal–Wallis one-way ANOVA test

Factor variable	Austin	Boston	Chicago	Los Angeles	Nashville
Superhost	73.197***	35.327***	17.67***	2.568,3#	56.222***
Host_Verified	9.950,9***	16.504***	15.371***	17.384***	12.016***
Property_Type	589.07***	451.96***	333.44***	1,394.6***	514***
Room_Type	3,013.5***	1,625.5***	3,198.2***	16,933***	1,321.5***
Booking	28.716***	18.916***	1.080,3***	62.538***	144.34***
Cancellation	502.49***	246.68***	180.24***	1,924.3***	597.79***
Guest_Picture	1.262,1#	114.96***	1.571,8#	20.912***	17.194***

Notes: # Not significant; ***significant at a 1% level of significance

PSO for tuning hyper-parameters of the ensemble ML algorithms. As constructing advanced ML models on a relatively large data set is highly sensitive to resource and computational time, a complete brute-force search to perform tuning of parameters will amplify the appetite for additional expenses. Please see Jana *et al.* (2021) for operational steps and details.

Gradient boosting

GB applies a bunch of base learners sequentially in the forward direction to produce the outcome (Schapire and Singer, 1999). It is a variation of the standard boosting methodology. Please see Schapire and Singer (1999) for details about the technique.

Bagging

Bagging is an ensemble predictive analytics approach. The outcome of the base learners (decision trees) constructed, applying bootstrapped samples, decides the final prediction (Simidjievski *et al.*, 2015). Bagging reduces the variation of volatile learning techniques for forecasting tasks, which leads to better prediction. For details steps of the bagging learning procedure, please see Simidjievski *et al.* (2015).

Random forest

RF is a typical ensemble ML algorithm (Breiman, 2001). Unlike GB, RF deploys parallel base learners to obtain the outcome. Likewise, GB and regression trees work as base learners. The final forecast is fetched by performing an arithmetic average of the individual outcome. Decision trees, grown on a randomly selected subset of the training segment, have been used as base learners. We see the successful application of these algorithms in the predictive modeling of financial time series (Ghosh *et al.*, 2018, 2019).

Stacking

It is an ensemble modeling framework. The method effectively treats the prediction of the different algorithms as input and the target variable as the output. In general, a separate learning algorithm, different from the predictive modeling techniques, generates inputs for stacking. We use XGBoost (Friedman, 2001), an extension of the boosting algorithm, to combine the forecasts found from GB, bagging and RF to generate final forecasts. These three ensemble ML algorithms are very effective for predictive analytics of extremely complex data sets (Lemmens and Croux, 2006; Ghosh *et al.*, 2019). The methodological approach for predicting rental prices of select Airbnb units is accomplished in two steps. In the first step, GB, bagging and RF are separately trained on the filtered set of features by the Boruta algorithm to estimate the rental prices. Parameter tuning of the three models is carried out using the PSO algorithm. In the next step, the predicted figures of the three models are combined in the stacking model by the XGBoost algorithm to derive the final predicted figures of the rental prices. The hyper-parameters of the XGBoost model are auto-tuned using the PSO algorithm.

Explainable artificial intelligence

We explain the black box typed ensemble ML-inspired predictive framework for drawing key implications and insights. Explainable AI, a useful ensemble ML technique, has been used. Recently, adoptions of Explainable AI in exploring and interpreting complex ML models have seen considerable traction in the literature (Jana *et al.*, 2022; Kaadoud *et al.*, 2022; Wang *et al.*, 2022). We use the "Shapash" library of Python, which has two

Explainable AI components – Shapley additive explanation (SHAP) (Lundberg and Lee, 2017) and local interpretable model-agnostic explanations (LIME) (Ribeiro *et al.*, 2016). Parallel to predictive analytics of Airbnb listing prices, Explainable AI caters to insights pertinent to influencing the structure of considered determinants. Mathematically, the SHAP values reflecting the feature importance are computed using equation (1):

$$\varnothing_i = \sum\nolimits_{S \subseteq N\{i\}} \frac{|S|!(n-|S|-1)!}{n!}\left[p(S \cup \{i\}) - p(S)\right] \tag{1}$$

in which $\varnothing_i$ signifies the impact of the ith feature, N denotes the features set having cardinality n, $p(N)$ is the forecasted figure for the ith feature and $S \in N$ with feature i. The feature explanation is written as:

$$h(y') = \varnothing_0 + \sum\nolimits_{j=1}^{F} \varnothing_j y_j' \tag{2}$$

in which $y' \in \{0,1\}^F$, F is the number of considered features.

The use of Explainable AI assists in uncovering the dependence interplay between the filtered set of explanatory features by the Boruta and Airbnb rental prices in US cities globally and locally. The process is shown in Figure 1.

Performance evaluation

We use four indices to evaluate the predictive performance.

Nash-Sutcliffe Efficiency (NSE): it is expressed as the ratio of residual variance obtained from a predictive model and the original variance of the data set. We compute NSE as follows:

$$NSE = 1 - \frac{\sum\nolimits_{t=1}^{N}\left\{\hat{Y}_t - Y_t\right\}^2}{\sum\nolimits_{i=1}^{N}\left\{Y_t - \overline{Y_t}\right\}^2} \tag{3}$$

where Y_t, $\overline{Y_t}$ and $\hat{Y}_t$ represents original data points, the average of original values and predicted figures. The NSE estimates range from $-\infty$ to 1. An estimate near to one implies highly accurate predictions.

Coefficient of determination (R^2): the measures R^2 is defined as:

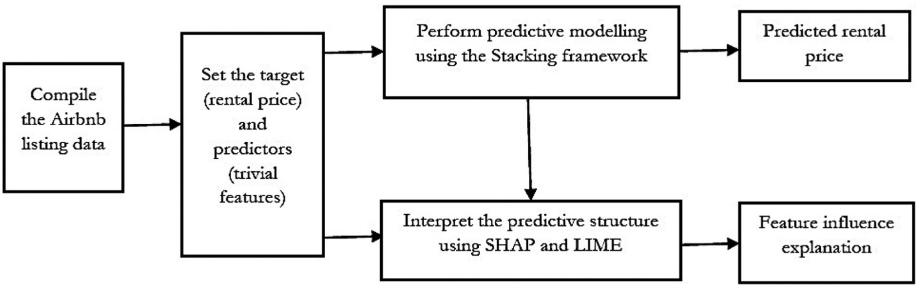

Figure 1.
Use of Explainable AI

$$R^2 = \frac{\sum_{t=1}^{N}\left\{\hat{Y}_t - \overline{Y_t}\right\}^2}{\sum_{i=1}^{N}\left\{Y_t - \overline{Y_t}\right\}^2} \tag{4}$$

The value of R^2 should ideally be close to one for an efficient model.

Index of agreement (IA): it gauges the model residual as:

$$IA = 1 - \frac{\sum_{t=1}^{N}\left(\hat{Y}_t - Y_t\right)^2}{\sum_{i=1}^{N}\left\{\left|\hat{Y}_t - \overline{Y_t}\right| + \left|Y_t - \overline{Y_t}\right|\right\}^2} \tag{5}$$

For the best prediction quality, IA values should be close to one.

Theil index (TI): we measure the TI as:

$$TI = \frac{\left[\frac{1}{N}\sum_{t=1}^{N}\left(\hat{Y}_t - Y_t\right)^2\right]^{1/2}}{\left[\frac{1}{N}\sum_{i=1}^{N}\left(\hat{Y}_t\right)^2\right]^{1/2} + \left[\frac{1}{N}\sum_{i=1}^{N}\left(Y_t\right)^2\right]^{1/2}} \tag{6}$$

TI must be sufficiently closer to zero for a proficient predictive model.

The above indicators have been effectively used for ascertaining the performance of predictive analytics (Ghosh *et al.*, 2022; Jana *et al.*, 2022). We also use superior predictive ability (SPA) and equal predictive ability statistical assessments for executing the task. The Diebold–Mariano (DM) test checks the equality of predictive capability, and the model confidence set (MCS) test ascertains SPA. If the DM test fails to distinguish the performance, we use the SPA.

Results and analyses

We present here the outcome of feature selection, predictive modeling and comparative performance evaluation.

Feature selection

The Boruta algorithm-based feature screening process is simulated using the "Boruta" library of R. We run the Boruta algorithm for 1,000 iterations to draw the final inferences. Table 2 reports the ranking of the 25 features related to Airbnb listing prices based on their explanatory capabilities. Boruta filters out irrelevant features by comparing performance with shadow attributes.

The outcome suggests that all the features are important for the Airbnb units in Austin. Few features are insignificant for rental price prediction in the remaining four cities. Features like Superhost, Property_Type, Review_Scores and Cancellation are insignificant for estimating listing prices of Airbnb units in Boston. Rental rates of Chicago Airbnb units are independent of Host_Verified, Minimum_Night, Review_Scores and Cancellation. Guest_Picture has no significant predictive power in Nashville. Los Angeles units are independent of seven generic features – Host_Verified, Property_Type, Maximum_Night, Review_Scores, Booking and Cancellation. The lack of predictive power of Host_Verified

Feature	Austin	Boston	Chicago	Los Angeles	Nashville
Superhost	7	25 (U)	19	14	22
Host_listings	2	1	5	11	2
Host_Verified	12	20	24 (U)	21 (U)	5
Property_Type	3	22 (U)	21	19 (U)	4
Room_Type	22	9	13	3	8
Accommodates	18	5	7	4	6
Bathrooms	16	10	11	1	3
Bedrooms	24	13	3	2	11
Beds	17	11	15	10	9
Guests	20	7	16	13	10
Extra_People	13	21	18	17	16
Minimum_Night	6	4	25 (U)	8	15
Maximum_Night	15	19	2	25 (U)	23
Availability_30	19	17	12	5	20
Availability_60	11	16	8	6	17
Availability_90	10	15	4	7	18
Availability_365	5	12	1	12	21
Reviews	4	8	10	15	14
Reviews_12	14	14	9	16	13
Review_Scores	23	24 (U)	22 (U)	18 (U)	24
Booking	21	18	20	20 (U)	19
Cancellation	9	23 (U)	23 (U)	22 (U)	7
Guest_Picture	25	2	16	24	25 (U)
Total_Listings	1	3	6	9	1
Review_M	8	6	14	23 (U)	12

Note: U = unimportant

Table 2.
Feature ranking by Boruta

and Cancellation indicates the insignificance of the credibility of hosts and the booking cancellation procedures in driving rental prices. It is essential to filter out the insignificant features beforehand, as their presence in the predictive framework may lead to overfitting issues. Thus, unimportant features do not appear for subsequent predictive modeling of respective cities. The ranking of features is not uniform across all five cities. So, the predictive capability of the features varies over the cities in estimating Airbnb listing prices. Spearman rank correlation statically verifies this claim. Table 3 reports the results.

The significance level of the test statistics further justifies the claim that independent features' explanatory power is not uniform. The Kolmogorov–Smirnov and Mann–Whitney statistical tests establish that the distribution of Airbnb listing prices in five cities is different. Similarly, the explanatory capabilities of the features are also nonuniform.

Cities	Austin	Boston	Chicago	Los Angeles	Nashville
Austin	–	0.07846#	0.09117#	−0.09231#	0.32538#
Boston		–	0.35122#	0.31308#	0.32615#
Chicago			–	0.39969#	0.00730#
Los Angeles				–	0.30385#
Nashville					–

Note: # Not significant

Table 3.
Outcome of the rank correlation test

Predictive modeling
The feature set selected by Boruta builds the predictive models. The “scikit-learn” library of Python simulates the four predictive models. The performance of the algorithms relies upon various process parameters. We use PSO to fine-tune the ensemble learning techniques for drawing superior predictions. The predictive structure has been implemented in Python programming through the customization of “sklearn” and “pspso” packages (Haider *et al.*, 2021). For example, the base learners’ numbers and the maximum features’ number for branching in base learners in all three methods have been auto-tuned using the PSO algorithm. A population swarm size of 50 and 500 iterations are fixed for simulating the search algorithm. Model-specific process parameters, namely, maximum depth in GB, bootstrap features in bagging, minimum samples in the leaf node, etc., have been fine-tuned. The PSO framework is also leveraged to build the final stacking framework, wherein the process parameters of the extreme GB algorithm are also optimized. Table 4 outlines the process parameters tuned by the PSO algorithm.

We segregate data sets into training (80%) and test (20%) subsets. Four performance indicators evaluate the predictive performance. We found that the IA, R^2 and NSE values are less than 0.95 for all cities except Boston. As a result, the predictive performance was not stubbornly excellent. Thus, the deployment of the stacking framework to improve the prediction quality is justified. Stacking then combines the output of three ensemble methods in an ensemble of ensemble framework to produce the final forecast, similarly applying PSO to auto-tune the backend algorithm, XGBoost, for implementation. Table 5 reports the stacking approach’s performance.

The performance indicators values improved considerably in both training and test data segments. From the performance measures values, we observe that the accuracy of forecasted listing prices of Airbnb units in Boston is superior, followed by Austin and Los Angeles. The rental prices in Chicago and Nashville turn out to be less predictable.

Thus, incorporating PSO for fine-tuning hyper-parameters of learning algorithms unearths the inherent patterns linking Airbnb listing prices with explanatory features. The role of initial feature screening through the Boruta algorithm is important. Discarding irrelevant features beforehand while proceeding with the predictive modeling to avert

Algorithms	Process parameters
Gradient boosting	{number of base learners: [500, 1,000, 1,500, 2,000, 2,500, 3,000], maximum features: [“auto,” “sqrt,” “log2,” 1, 2, 4, 6, 8, 10], learning rate: [0.010, 0.015, 0.020, 0.025, 0.030, 0.035, 0.040, 0.045, 0.050, 0.055, 0.060, 0.065, 0.070, 0.075, 0.080, 0.085, 0.090, 0.095, 0.1], maximum depth: [2, 3, 4, 5, 6, 7, 8, 9, 10]}
Bagging	{number of base learners: [500, 1,000, 1,500, 2,000, 2,500, 3,000], maximum features: [“auto,” “sqrt,” “log2” 1, 2, 3, 4, 5, 6, 7, 8, 9, 10], bootstrap features: [“True,” “False”], warm start: [“True,” “False”]}
Random forest	{number of base learners: [500, 1,000, 1,500, 2,000, 2,500, 3,000], maximum features: [“auto,” “sqrt,” “log2,” 1, 2, 3, 4, 5, 6, 7, 8, 9, 10], minimum samples for splitting: [1, 2, 3, 4, 5, 6, 7, 8, 9, 10], minimum samples in leaf node: [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]}
Stacking (extreme gradient boosting)	{number of base learners: [500, 1,000, 1,500, 2,000, 2,500, 3,000], learning rate: [0.010, 0.015, 0.020, 0.025, 0.030, 0.035, 0.040, 0.045, 0.050, 0.055, 0.060, 0.065, 0.070, 0.075, 0.080, 0.085, 0.090, 0.095, 0.1], maximum depth: [2, 3, 4, 5, 6, 7, 8, 9, 10]}

Table 4.
Parameter levels considered for tuning

overfitting problems is important. Stacking has emerged as a viable option for combining ensemble modeling and generating predictions mimicking an ensemble of an ensemble approach. Next, we conduct a comparative statistical assessment of the predictive performance of respective models, as it is necessary to check whether the combination of ensemble algorithms in stacking generates statistically superior forecasts. The DM test assesses the comparative assessment. We use indexes for denoting the model number, as the rest is performed pairwisely. Table 6 reports the outcome of the tests.

Table 5. Predictive performance of stacking

Performance indicators	Austin	Boston	Chicago	Los Angeles	Nashville
Training data set					
IA	0.9915	0.9978	0.9822	0.9853	0.9837
TI	0.0216	0.0198	0.0239	0.0233	0.0229
R^2	0.9862	0.9954	0.9741	0.9768	0.9760
NSE	0.9838	0.9933	0.9736	0.9759	0.9751
Test data set					
IA	0.9866	0.9951	0.9776	0.9796	0.9781
TI	0.0228	0.0204	0.0247	0.0238	0.0242
R^2	0.9623	0.9939	0.9712	0.9739	0.9723
NSE	0.9810	0.9925	0.9690	0.9714	0.9706

Table 6. Outcome of the DM test comparative analysis

Models	Gradient boosting (1)	Bagging (1)	Random forest (1)	Stacking (1)
		Austin		
Gradient boosting (2)	–			
Bagging (2)	0.241#	–		
Random forest (2)	0.229#	0.245#	–	
Stacking (2)	4.9617***	5.0238***	4.9876***	–
		Boston		
Gradient boosting (2)	–			
Bagging (2)	0.223#	–		
Random forest (2)	0.235#	0.231#	–	
Stacking (2)	4.9295***	5.0166***	4.9853***	–
		Chicago		
Gradient boosting (2)	–			
Bagging (2)	0.233#	–		
Random forest (2)	0.226#	0.215#	–	
Stacking (2)	5.8761***	5.796***	5.8204***	–
		Los Angeles		
Gradient boosting (2)	–			
Bagging (2)	0.223#	–		
Random forest (2)	0.236#	0.221#	–	
Stacking (2)	5.8432***	5.8138***	5.7689***	–
		Nashville		
Gradient boosting (2)	–			
Bagging (2)	0.231#	–		
Random forest (2)	0.234#	0.227#	–	
Stacking (2)	5.8530***	5.8688***	5.8726***	–

Notes: # Not significant, ***significant at a 1% level of significance

Predictions from the stacking approach are statistically superior to the other three models for all five cities. Thus, the proposed ensemble of ensembles approach has successfully enhanced the quality of the final prediction in comparison to standalone ensemble algorithms. Therefore, we can effectively predict the listing prices of any Airbnb units using the presented research framework based on the actual offerings of the given set of features. We can also predict Airbnb units of which city is more predictable. A comparison of city-wise predictability is critical to judge for meaningful implications. However, the mere usage of the DM test for equal predictability assessment failed to discriminate between Austin, Boston and Los Angeles as statistically more predictable than Chicago and Nashville. It also failed to break the tie between Chicago and Nashville to comprehend, which is the least predictable. Thus, it is challenging to identify the most and the least predictable cities using the test of equal predictability. Hence, we use the MCS test to check the SPA. The outcome of the MCS test suggests that the listing prices of Airbnb units in Boston are most predictable, followed by Austin and Chicago. According to the descriptive statistics, we can recall that Boston’s unit prices have less variation. Nashville and Chicago took the 4th and 5th spots, indicating the predictive performance is not as good as the other three cities.

A comparative analysis using MCS evaluation rationalizes the efficiency and superiority of Airbnb listing price prediction. We further compare the proposed framework’s predictive performance with the following variants of ML techniques: proposed stacking framework without parameter tuning by PSO (Stacking-WoP); support vector regression (SVR), a conventional ML model; regularized random forest (RRF), a stand-alone ensemble technique; and categorical boosting (CatBoost), a variant of GB. The SVR, RRF and CatBoost are used in two setups, tuned with PSO and without PSO, SVR-WoP, RRF-WoP and CatBoost_WoP. The comparative analysis is balanced and critically evaluates the contribution of PSO and ensemble learning approaches in augmenting the accuracy of the predictions. We perform ten experimental trials to identify the best possible combination of parameters in these competing models. In addition, we have not used the Boruta feature screening algorithm. Thus, the four competing models yield predictions considering all 25 features. Table 7 reports the outcome for respective cities.

Table 7 shows that the proposed stacking framework optimized through PSO is statistically superior. The utility of the Boruta algorithm is exemplified. The exclusion of PSO reduces the accuracy of predictions of the stacking process. SVR-WoP is the least effective in estimating Airbnb listing prices across the cities, whereas the PSO-tuned RRF and the CatBoost resemble the 6th and 7th spots. The efficiency of these methods in fetching accurate predictions is enhanced when auto-tuned with PSO, as apparent from the relative rankings. Hence, the true utility of the PSO in the integrated predictive structure is justified. Stacking-WoP has secured the 2nd spot in predictive accuracy, emphasizing the advantage of the combination of ensemble ML over the stand-alone approaches. Therefore, the inclusion of PSO and stacking-based ensemble framework is highly profound and reliable in uncovering the dynamics of Airbnb rental prices. The proposed integrated predictive

City	Stacking	Stacking-WoP	SVR-WoP	RRF-WoP	CatBoost-WoP	SVR	RRF	CatBoost
Austin	(1)	(2)	(8)	(6)	(7)	(5)	(3)	(4)
Boston	(1)	(2)	(8)	(7)	(6)	(5)	(4)	(3)
Chicago	(1)	(2)	(8)	(7)	(6)	(5)	(4)	(3)
Los Angeles	(1)	(2)	(8)	(6)	(7)	(5)	(3)	(4)
Nashville	(1)	(2)	(8)	(7)	(6)	(5)	(4)	(3)

Table 7. Outcome of comparative predictive evaluation

modeling framework has accomplished the research endeavors, as the predicted figures' quality is apparent. Thus, this study contributes to voiding the existing research gaps. The proposed model can predict the Airbnb listing prices with supreme accuracy without considering the plethora of amenities pertinent to the various appliance. Among all the cities, Airbnb listing prices in Boston have the highest degree of predictability. The proposed stacking structure that combines three ensemble ML algorithms survives the critical statistical checks. Therefore, the framework estimates accurate forecasts and is deemed to be classified as robust. Its statistical superiority over the individual ensemble ML models and the benchmark predictive tools, as apparent from the outcome of the DM and MCS tests, justifies the methodological contribution.

Model interpretation

After completing the predictive analysis of large-scale data, we explain the critical insights on feature influence. Figures 2–3 exhibit the output of the Explainable AI modeling of Airbnb listings in two sample locations, Austin and Boston and Chicago, to comprehend the nature of the contribution of the explanatory features. The global influence patterns of the 20 best explanatory variables are shown in the top left corner exhibit. The variations of the predictive contribution of the explanatory variables are reflected in the top right corner

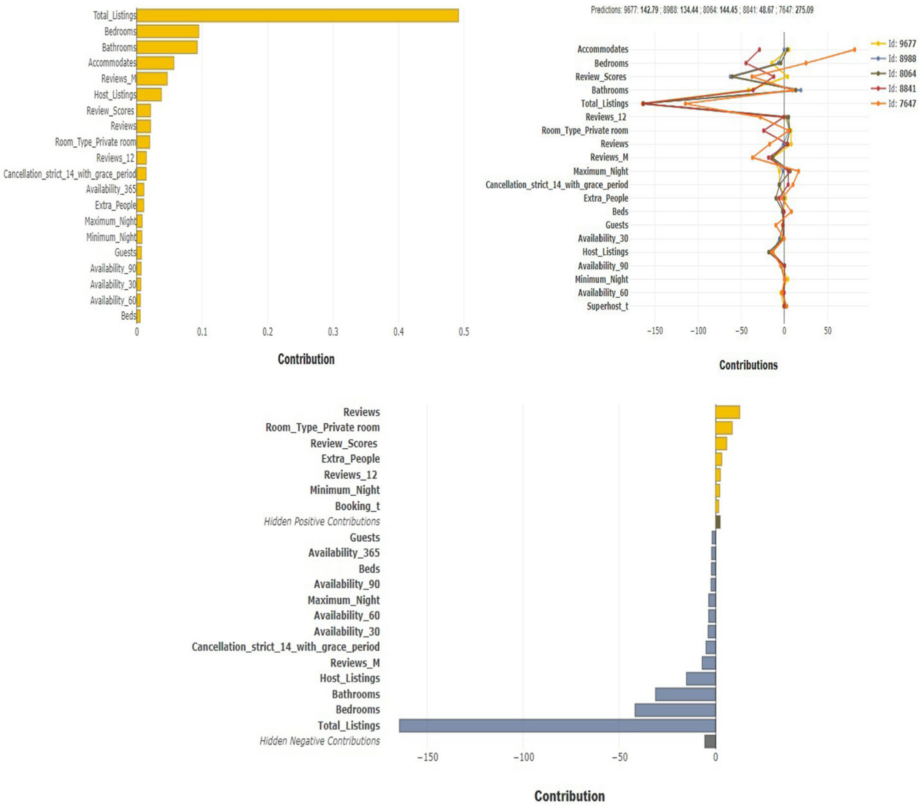

Figure 2.
Outcome of Explainable AI for modeling Austin units

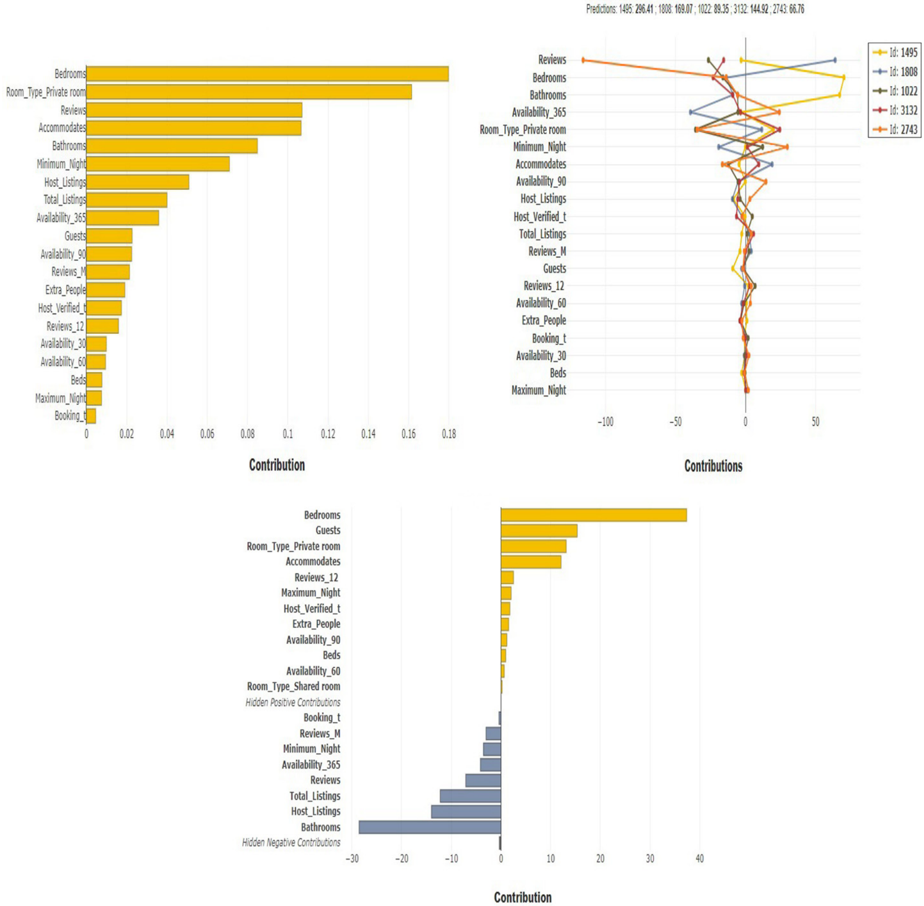

Figure 3.
Outcome of Explainable AI for modeling Boston units

exhibit. Finally, the bottom exhibit depicts the influence structure locally on a random sample as determined by the LIME framework.

We find that Total_Listings is the most influential variable for Airbnb units in Austin. Features like bedrooms, bathrooms, etc., are critical in driving the listing prices and are apparent from the top right corner figure. The outcome of LIME-based introspection of features' influence at the local level hints at variation in the strength and direction of influence of the underlying features. Some features are likely to negatively impact listing prices at the local scale. Hence, incorporating all filtered features through the Boruta algorithm is of paramount significance for achieving precise predictions at the local level.

For Airbnb units in Boston, the number of bedrooms is the most important driver of listing prices. On the other hand, accommodation in private rooms and bathrooms has resembled 2nd and 3rd spots in feature importance ranking. Also, the influence of local-level variations and different directions of impact is visible.

Detailed results of model interpretation in other locations are available on request from the authors. Feature contribution ranking across Chicago and Los Angeles Airbnb units are identical, wherein bedrooms largely explained the variation. Reviews by guests, bathrooms, etc., are influential features for the Nashville units. The local-level fluctuations of rental prices in these three places could be tracked by monitoring privacy and flexibility to accommodate extra people-related factors. In general, it has been found that the features like bedrooms and bathrooms are the dominant ones driving listing prices of the considered data set. Information pertinent to listings, reviews, etc., also provides relatable insights. Monitoring all screened features used in building predictive models is necessary, as achieving high accuracy across all samples occasionally depends on profound contributions from apparently unimportant features.

Conclusions

This research develops an integrated nature-inspired optimization governed ensemble learning-based predictive modeling framework. The findings indicate the efficacy of the proposed approach in accomplishing essential research endeavors. The framework's scalability and implementation ease in discovering the hidden patterns of large voluminous data have been established. The proposed approach used a common set of features for all accommodation units and derived predictions. The model can effectively be used by hosts to fix realistic listing prices. Statistical analyses of the overall findings indicate that the present research fills the existing research gaps. The inclusion of Explainable AI has contributed to P2P accommodation literature. The existing literature is replete with hedonic regression-driven models to quantify the impact of various amenity and environmental factors. These models are not ideal for yielding highly accurate predictions. The accuracy of the prediction model is more important for a host to precisely estimate the revenue margins. The present work elucidates the efficacy and reliability of the proposed framework in fetching superior predictions. The model interpretation facilitates the identification of the major drivers to understand customer preferences. Hence, the utility of the robust predictive model and subsequent model explanation can be considered a substantial contribution to existing research. The present research espouses the utility of AI and advanced tools for the betterment of holistic management of the tourism and hospitality sector (Ampountolas and Legg, 2021; Elkhwesky and Elkhwesky, 2022). The significant outcomes of this research are outlined below:

The independent features do not uniformly influence the listing prices of Airbnb units in different locations. The explanatory power of features varied over the places. The feature set used in this research assists to a large extent in obtaining precise forecasts. Boruta feature selection algorithm successfully identifies city-wise most significant features. A large number of amenity-driven features are not necessary to incorporate for predicting listing prices. Airbnb listings in Boston are the most predictable, followed by Austin and Los Angeles. The listing prices of Chicago Airbnb units are the least predictable. The highly volatile nature of Chicago and Nashville Airbnb units needs to be monitored appropriately. The proposed stacking architecture incorporating PSO as parameter tuning yields statistically superior forecasts of Airbnb listing prices across the cities. Generic features like Bedrooms, Bathrooms, Total_Listings and Accommodates possess relatively more predictive power in explaining listing prices. The key drivers of Airbnb listing prices in Chicago and Los Angeles are similar.

Theoretical implications

Seamless integration of feature screening through Boruta and nature-inspired optimized ensemble ML has proved efficient in modeling the rental pricing dynamics of the largest sharing economy platform. The framework is highly effective for the tourism sector as proper usage of AI acts as a major differentiator (Filieri *et al.*, 2021). The literature is replete with advanced data modeling frameworks for empirical analysis of the travel, tourism and hospitality sectors (Kumar *et al.*, 2019; Jana and Mitra, 2021). The addition of the Explainable AI components has resulted in several interesting findings. Bedrooms, Bathrooms, Total_Listings, Accommodates, etc., are the most dominant features in driving the Airbnb listing prices. Thus, from the host's perspective, it would be meaningful to emphasize these features carefully to augment customer experience, eventually resulting in higher revenue in the long run. The listing price dynamics of Chicago and Los Angeles are almost identical through the lens of Explainable AI. A deeper investigation is required to infer whether any particular trait of consumer behavioral patterns is interlinked with the said phenomenon.

Practical implications

The outcome of Explainable AI indicates the willingness of travelers to pay a price premium for privacy and luxurious offerings in the chosen locations. The accuracy of predictions in the absence of specific spatial features implies that offering-related attributes primarily drive the sharing economy market in the USA. This phenomenon suggests the dominance and capacity of the sharing economy platform in sweeping the travel and hospitality sector, inhibiting hazardous external factors. In the context of the overall tourism sector growth, the surge in demand for privacy and other allied facilities can daunt conventional economic hotels and other cheap offerings that obscure customer preferences. Hosts can work on the revenue margin by focusing on crucial influential generic features accordingly in their respective locations. The interpretable framework has practical implications for strategizing pricing for sustainability in the long run. Unlike the orthodox hedonic regression models for determining price based on specific amenities, environmental factors, etc., the proposed framework presents a scalable approach considering a common set of features to precisely predict the rental prices of one of the largest sharing economy platforms in the USA. The feature ranking enables the hosts to look for alternative offerings to compensate for the revenue if investing in the prime features is beyond the budget. Accordingly, paying mortgages can be planned, if any. Therefore, the present study offers strategic roadmaps for new and existing hosts to sustain in the competing sharing economy platform. End users can screen out the rental offerings in respective cities predominantly based on essential features like the number of bedrooms, availability of the private room, bathrooms, etc., as per the budget.

Limitations and future research

We obtained the present study results based on data sets from five US cities. Application of the approach to Airbnb units in other countries will be interesting and carrying out an exhaustive comparative study. It would be challenging and important to gauge the impact of geographical factors, the extent of urbanization and the natural environment in governing the predictability of Airbnb rental prices in a cross-country context. We have used November 2019 data sets to avoid the destruction caused by the COVID-19 pandemic. Analyzing the spatial and temporal variations of rental prices explicitly during the pandemic will be a worthy future research agenda. It will be extremely arduous and challenging to identify the appropriate variables to tame the impact of an uncertain external environment while comprehending the price dynamics in the new-normal time. We can combine text mining models with the presented predictive framework for analyzing the sentiment expressed in customer reviews to further augment the predictions' quality. The present study shows the efficacy of ensemble modeling. It will be interesting to check the efficiency of deep learning algorithms in estimating Airbnb prices and conducting comparative research.

References

Ampountolas, A. and Legg, M.P. (2021), "A segmented machine learning modeling approach of social media for predicting occupancy", *International Journal of Contemporary Hospitality Management*, Vol. 33, pp. 2001-2021.

Breiman, L. (2001), "Random forests", *Machine Learning*, Vol. 45 No. 1, pp. 5-32.

Chica-Olmo, J., González-Morales, J.G. and Zafra-Gómez, J.L. (2020), "Effects of location on Airbnb apartment pricing in Málaga", *Tourism Management*, Vol. 77, p. 103981.

Doborjeh, Z., Hemmington, N., Doborjeh, M. and Kasabov, N. (2022), "Artificial intelligence: a systematic review of methods and applications in hospitality and tourism", *International Journal of Contemporary Hospitality Management*, Vol. 34, pp. 1154-1176.

Dogru, T., Majid, K., Laroche, M., Mody, M. and Suess, C. (2021), "Communicating quality while evoking loss – how consumers assess extra charges in the Airbnb marketplace", *Tourism Management*, Vol. 87, p. 104376.

Elkhwesky, Z. and Elkhwesky, E.F.Y. (2022), "A systematic and critical review of internet of things in contemporary hospitality: a roadmap and avenues for future research", *International Journal of Contemporary Hospitality Management*, Vol. 35, pp. 533-562, doi: 10.1108/IJCHM-01-2022-0090.

Farmaki, A., Spanou, E. and Christou, P. (2021), "Constructing meanings of luxury in Airbnb", *International Journal of Contemporary Hospitality Management*, Vol. 33, pp. 3306-3330.

Filieri, R., D'Amico, E., Destefanis, A., Paolucci, E. and Raguseo, E. (2021), "Artificial intelligence (AI) for tourism: an European-based study on successful AI tourism start-ups", *International Journal of Contemporary Hospitality Management*, Vol. 33, pp. 4099-4125.

Friedman, J.H. (2001), "Greedy function approximation: a gradient boosting machine", *Annals of Statistics*, Vol. 29, pp. 1189-1232.

Ghosh, I., Datta Chaudhuri, T., Alfaro-Cortés, E., Gámez, M. and García, N. (2022), "A hybrid approach to forecasting futures prices with simultaneous consideration of optimality in ensemble feature selection and advanced artificial intelligence", *Technological Forecasting and Social Change*, Vol. 181, p. 121757.

Ghosh, I., Jana, R.K. and Sanyal, M.K. (2019), "Analysis of temporal pattern, causal interaction and predictive modeling of financial markets using nonlinear dynamics, econometric models and machine learning algorithms", *Applied Soft Computing*, Vol. 82, p. 105553.

Ghosh, I., Sanyal, M.K. and Jana, R.K. (2018), "Fractal inspection and machine learning-based predictive modelling framework for financial markets", *Arabian Journal for Science and Engineering*, Vol. 43 No. 8, pp. 4273-4287.

Guttentag, D. (2015), "Airbnb: disruptive innovation and the rise of an informal tourism accommodation sector", *Current Issues in Tourism*, Vol. 18, pp. 1192-1217.

Gyodi, K. and Nawaro, L. (2021), "Determinants of Airbnb prices in European cities: a spatial econometrics approach", *Tourism Management*, Vol. 86, p. 104319.

Haider, A., Field, M., Sykes, J., Carolan, M. and Holloway, L. (2021), "PSPSO: a package for parameters selection using particle swarm optimization", *Software X*, Vol. 15, p. 100706.

Han, W. and Bai, B. (2022), "Pricing research in hospitality and tourism and marketing literature: a systematic review and research agenda", *International Journal of Contemporary Hospitality Management*, Vol. 34, pp. 1717-1738.

Hu, L., He, S., Han, Z., Xiao, H., Su, S., Weng, M. and Cai, Z. (2019), "Monitoring housing rental prices based on social media: an integrated approach of machine-learning algorithms and hedonic modeling to inform equitable housing policies", *Land Use Policy*, Vol. 82, pp. 657-673.

Islam, M.D., Li, B., Islam, K.S., Ahasan, R., Mia, M.R. and Haque, M.E. (2022), "Airbnb rental price modeling based on latent Dirichlet allocation and MESF-XGBoost composite model", *Machine Learning with Applications*, Vol. 7, p. 100208.

Jana, R.K., Ghosh, I. and Das, D. (2021), "A differential evolution-based regression framework for forecasting bitcoin price", *Annals of Operations Research*, Vol. 306 No. 1, pp. 295-320.

Jana, R.K., Ghosh, I. and Sanyal, M.K. (2020), "A granular deep learning approach for predicting energy consumption", *Applied Soft Computing*, Vol. 89, p. 106091.

Jana, R.K., Ghosh, I. and Wallin, M.W. (2022), "Taming energy and electronic waste generation in bitcoin mining: insights from Facebook prophet and deep neural network", *Technological Forecasting and Social Change*, Vol. 178, p. 121584.

Jana, R.K. and Mitra, S.K. (2021), "Maximizing pleasure in day tours: utility theory approach", *International Journal of Tourism Research*, Vol. 23 No. 4, pp. 532-541.

Kaadoud, I.C., Bennetot, A., Mawhin, B., Charisi, V. and Díaz-Rodríguez, N. (2022), "Explaining aha! Moments in artificial agents through IKE-XAI: implicit knowledge extraction for eXplainable AI", *Neural Networks*, Vol. 155, pp. 95-118, doi: 10.1016/j.neunet.2022.08.002.

Kalehbasti, P.R., Nikolenko, L. and Rezaei, H. (2021), "Airbnb price prediction using machine learning and sentiment analysis", in Holzinger, A., Kieseberg, P., Tjoa, A.M., Weippl, E. (Eds), *Machine Learning and Knowledge Extraction. CD-MAKE 2021*, Lecture Notes in Computer Science, Springer, Cham, Vol. 12844.

Kennedy, J. and Eberhart, R. (1995), "Particle swarm optimization", *Proceedings of ICNN'95 – International Conference on Neural Networks*, pp. 1942-1948, doi: 10.1109/ICNN.1995.488968.

Kumar, M., Prashar, S. and Jana, R.K. (2019), "Does international tourism spur international trade and output? Evidence from wavelet analysis", *Tourism Economics*, Vol. 25 No. 1, pp. 22-33.

Kursa, M.B. and Rudnicki, W.R. (2010), "Feature selection with the Boruta package", *Journal of Statistical Software*, Vol. 36, pp. 1-13.

Lagonigro, R., Martori, J.C. and Apparicio, P. (2020), "Understanding Airbnb spatial distribution in a Southern European city: the case of Barcelona", *Applied Geography*, Vol. 115, p. 102136.

Lawani, A., Reed, M.R., Mark, T. and Zheng, Y. (2019), "Reviews and price on online platforms: evidence from sentiment analysis of Airbnb reviews in Boston", *Regional Science and Urban Economics*, Vol. 75, pp. 22-34.

Lee, C.K.H., Tse, Y.K., Zhang, M. and Wang, Y. (2022), "What have hosts overlooked for improving stay experience in accommodation-sharing? Empirical evidence from Airbnb customer reviews", *International Journal of Contemporary Hospitality Management*, Vol. 35, pp. 765-784, doi: 10.1108/IJCHM-12-2021-1544.

Lemmens, A. and Croux, C. (2006), "Bagging and boosting classification trees to predict churn", *Journal of Marketing Research*, Vol. 43, pp. 276-286.

Liu, Y. (2021), "Airbnb pricing based on statistical machine learning models", *IEEE International Conference on Signal Processing and Machine Learning (CONF-SPML)*, pp. 175-185, doi: 10.1109/CONF-SPML54095.2021.00042.

Liu, X., Zeng, Y., He, J. and Li, Z. (2022), "Value co-creation research in tourism and hospitality: a comparative bibliometric analysis", *International Journal of Contemporary Hospitality Management*, Vol. 34, pp. 663-686.

Lundberg, S. and Lee, S.I. (2017), "A unified approach to interpreting model predictions", arXiv: 1705.07874.

Mariani, M. and Baggio, R. (2022), "Big data and analytics in hospitality and tourism: a systematic literature review", *International Journal of Contemporary Hospitality Management*, Vol. 1, pp. 231-278.

Pradhan, A., Bisoy, S.K. and Das, A. (2021), "A survey on PSO based meta-heuristic scheduling mechanism in cloud computing environment", *Journal of King Saud University – Computer and Information Sciences*, Vol. 34, pp. 4888-4901, doi: 10.1016/j.jksuci.2021.01.003.

Qiu, H., Chen, D., Bi, J.-W., Lyu, J. and Li, Q. (2022), "The construction of the affinity-seeking strategies of Airbnb homestay hosts", *International Journal of Contemporary Hospitality Management*, Vol. 34, pp. 861-884.

Ribeiro, M.T., Singh, S. and Guestrin, C. (2016), "'Why should I trust you?' explaining the predictions of any classifier", *KDD 2016 San Francisco, CA, USA*, doi: 10.1145/2939672.2939778.

Sainaghi, R. (2021), "Determinants of price and revenue for peer-to-peer hosts. The state of the art", *International Journal of Contemporary Hospitality Management*, Vol. 33, pp. 557-586.

Sainaghi, R., Abrate, G. and Mauri, A. (2021), "Price and RevPAR determinants of Airbnb listings: convergent and divergent evidence", *International Journal of Hospitality Management*, Vol. 92, p. 102709.

Schapire, R.E. and Singer, Y. (1999), "Improved boosting algorithms using confidence-rated predictions", *Machine Learning*, Vol. 37, pp. 297-336.

Simidjievski, N., Todorovski, L. and Dzeroski, S. (2015), "Predicting long-term population dynamics with bagging and boosting of process-based models", *Expert Systems with Applications*, Vol. 42, pp. 8484-8496.

So, K.K.F., Kim, H. and Min, S. (2022), "Creating customer value in the sharing economy: an investigation of Airbnb users and their topographic characteristics", *International Journal of Contemporary Hospitality Management*, Vol. 34, pp. 23-45.

Song, X., Gu, H., Li, Y. and Ye, W. (2022), "A systematic review of trust in sharing accommodation: progress and prospects from the multistakeholder perspective", *International Journal of Contemporary Hospitality Management*, doi: 10.1108/IJCHM-12-2021-1555.

Thakur, N., Jain, R., Mahajan, A. and Islam, S.M.N. (2022), "Deep neural network based data analysis and price prediction framework for Rio De Janeiro Airbnb", *IEEE 7th International Conference for Convergence in Technology (I2CT)*, pp. 1-7, doi: 10.1109/I2CT54291.2022.9824383.

Tussyadiah, I.P. (2015), "An exploratory study on drivers and deterrents of collaborative consumption in travel", *In Information and Communication Technologies in Tourism 2015*, Springer, Cham, pp. 817-830.

Wang, D. and Nicolau, J.L. (2017), "Price determinants of sharing economy based accommodation rental: a study of listings from 33 cities on Airbnb.com", *International Journal of Hospitality Management*, Vol. 62, pp. 120-131.

Wang, R. and Rasouli, S. (2022), "Contribution of streetscape features to the hedonic pricing model using geographically weighted regression: evidence from Amsterdam", *Tourism Management*, Vol. 91, p. 104523.

Wang, Y., Liu, W. and Liu, X. (2022), "Explainable AI techniques with application to NBA gameplay prediction", *Neurocomputing*, Vol. 483, pp. 59-71.

Yang, S. (2021), "Learning-based Airbnb price prediction model", *2nd IEEE International Conference on E-Commerce and Internet Technology (ECIT)*, pp. 283-288, doi: 10.1109/ECIT52743.2021.00068.

Yu, C., Liao, W. and Margolin, D. (2021), "The differential impacts of blinded online reviews: comparing socio-emotional features of guest and host reviews on Airbnb", *Telematics and Informatics*, Vol. 65, p. 101731.

Zhang, L., Yan, Q. and Zhang, L. (2022), "A text analytics framework for understanding the relationships among host self-description, trust perception and purchase behavior on Airbnb", *Decision Support Systems*, Vol. 133, p. 113288.

Zhu, J., Cheng, M. and Li, Z. (2021), "Peer-to-peer accommodation experience and guest actual recommendations: a novel mixed-method approach", *Tourism Management Perspectives*, Vol. 38, p. 100816.

Zhu, L., Cheng, M. and Wong, I.A. (2019), "Determinants of peer-to-peer rental rating scores: the case of Airbnb", *International Journal of Contemporary Hospitality Management*, Vol. 31, pp. 3702-3721.

Further reading

Barnes, S.J. and Kirshner, S.N. (2021), "Understanding the impact of host facial characteristics on Airbnb pricing: integrating facial image analytics into tourism research", *Tourism Management*, Vol. 83, p. 104235.

Chang, H.H. (2020), "Does the room sharing business model disrupt housing markets? Empirical evidence of Airbnb in Taiwan", *Journal of Housing Economics*, Vol. 49, p. 101706.

Chen, C.C. and Chang, Y.C. (2018), "What drives purchase intention on Airbnb? Perspectives of consumer reviews, information quality, and media richness", *Telematics and Informatics*, Vol. 35 No. 5, pp. 1512-1523.

Corresponding author
Mohammad Zoynul Abedin can be contacted at: m.abedin@tees.ac.uk