E-ISSN: 2708-4523
P-ISSN: 2708-4515
Impact Factor (RJIF): 5.61
AJMC 2025; SP-6(2): 338-344

www.allcommercejournal.com
Received: 13-06-2025
Accepted: 15-07-2025

Abburi Sravani
2nd Year, M.S Data Science EdTech Division_Exafluence INC, Sri Venkateswara University Tirupati, Andhra Pradesh, India

Leela Krishna R
2nd Year, M.S Data Science EdTech Division_Exafluence INC, Sri Venkateswara University Tirupati, Andhra Pradesh, India

Pravallika Rajyalakshmi C
2nd Year, M.S Data Science EdTech Division_Exafluence INC, Sri Venkateswara University Tirupati, Andhra Pradesh, India

Bharghav Madhiraju
Instructor, M.S Data Science EdTech Division_Exafluence INC, Sri Venkateswara University Tirupati, Andhra Pradesh, India

Corresponding Author:
Abburi Sravani
2nd Year, M.S Data Science EdTech Division_Exafluence INC, Sri Venkateswara University Tirupati, Andhra Pradesh, India

From data to delight: Predictive analytics for optimizing hospitality service ratings

Abburi Sravani, Leela Krishna R, Pravallika Rajyalakshmi C and Bharghav Madhiraju

DOI: https://www.doi.org/10.22271/27084515.2025.v6.i2Sc.741

Abstract
As the hospitality industry continues to evolve, machine learning plays a crucial role in improving guest satisfaction and operational efficiency. This research introduces a predictive model specifically designed for homestay businesses, utilizing supervised learning methods-namely decision trees and logistic regression-to anticipate essential rating factors such as cleanliness, communication, accuracy, and check-in experience. The data, collected from a start-up, includes entries with features related to booking methods, customer profiles, and rating scores. Data preprocessing and exploratory analysis are conducted using Python, followed by model training and evaluation to determine guest satisfaction drivers, including factors like stay month and booking channel. The models are validated using accuracy metrics to ensure their reliability and effectiveness in real- world scenarios. Beyond prediction, the study proposes incorporating an AI-based recommendation engine in the future to suggest the most suitable homestay options based on user history and preferences. This combined strategy- rating prediction and personalized recommendations-aims to improve service quality, uncover areas for enhancement, and promote long- term growth in the homestay sector. The flexible design of this framework ensures adaptability across similar businesses, offering scalable, data- driven solutions and providing valuable insights for decision-makers focused on personalized experiences and evidence-based service upgrades.

Keywords: Machine learning, homestay businesses, predictive model, guest satisfaction, supervised learning, recommendation engine

Introduction

The hospitality sector is undergoing a rapid transformation, largely influenced by rising guest expectations, growing data availability, and the integration of advanced technologies. As digital platforms become the primary medium for travelers to explore accommodation options, the focus has shifted from basic listings to curated, personalized experiences. In this evolving environment, machine learning (ML) and artificial intelligence (AI) are proving to be powerful tools for interpreting customer behavior, enhancing service quality, and supporting data-driven decision-making.

This study centers on developing a predictive model tailored for homestays and hotel-style flats, employing supervised learning algorithms. The primary aim is to leverage historical customer data-including booking preferences, duration of stay, room allocation, and service feedback-to forecast future ratings in critical service areas such as cleanliness, communication, accuracy, and overall value.

In the current phase of implementation, two machine learning algorithms-Decision Tree Classifier and Logistic Regression-have been successfully used to build predictive models. These tools enable business owners to identify and act upon potential shortcomings in service delivery. For example, patterns may reveal that certain rooms regularly receive lower cleanliness ratings or that bookings through specific channels lead to poorer communication feedback. By identifying such trends early, businesses can make targeted improvements to their services.

Looking forward, the project envisions expanding this predictive framework into a fully adaptive AI-based recommendation engine. This next phase aims toenhance the user experience by offering dynamic, personalized accommodation suggestions informed by historical guest behavior, preferences, and review insights.

System Description

Current Phase: Machine Learning-Based Rating Prediction

The present system is built upon a structured dataset obtained from a startup enterprise, including both user-input and system-generated information. After performing essential data preprocessing-such as managing missing values, encoding categorical variables, and normalizing rating fields-the following models were developed:

- **Decision Tree Classifier:** Used to categorize service aspects such as cleanliness, communication, and accuracy based on inputs like room assignment, month of stay, and booking channel.
- **Logistic Regression:** Utilized for predicting the “Value for Money” score, with influencing factors including the price paid and booking details.

These models were trained and validated using a standard train-test split method. Their performance ranged between 57% and 83% accuracy, indicating a reasonable predictive capability despite limitations in dataset size. Exploratory Data Analysis (EDA) tools- such as boxplots, violin plots, and pivot tables-were applied to visualize relationships and confirm hypotheses.

Upcoming Phase: AI-Powered Recommendation Engine

To increase the system’s effectiveness, the project’s next objective is to integrate an AI-driven recommendation module. This feature will go beyond static prediction by actively learning from user interactions and generating personalized accommodation suggestions. It will assess various behavioral signals, such as browsing habits, review sentiments, and prior bookings, to offer recommendations most relevant to each user.

Key techniques planned for implementation include:

- **Collaborative Filtering:** Suggests properties based on the preferences of users with similar patterns.
- **Content-Based Filtering:** Recommends listings with similar attributes to those that the user has previously favored.
- **Natural Language Processing (NLP):** Analyzes guest reviews to extract sentiment and relevant context to enrich recommendation quality.
- **Matrix Factorization (e.g., SVD/ALS):** Discovers hidden user-item relationships to enhance recommendation accuracy.

This AI-enhanced system will continuously evolve, learning from each interaction to refine its suggestions. The anticipated outcomes include improved customer satisfaction, increased bookings, and smarter operational decisions for service providers.

This project outlines a dual-phase solution to modernize service delivery in the hospitality sector. In its initial phase, supervised learning techniques have been effectively employed to predict guest ratings, providing actionable feedback to homestay businesses. The proposed second phase introduces an AI-based recommendation system designed to deliver tailored accommodation options, leveraging deeper insights from behavioral data.

Together, these components form an end-to-end intelligent analytics pipeline, enabling homestay providers to evolve from reactive service models to proactive, personalized guest experiences-setting a new standard in customer-centric hospitality.

Literature Review

Several research studies have explored how machine learning and data-driven techniques can be applied to improve services in the hospitality industry. This section highlights five key works relevant to rating prediction and guest sentiment analysis.

This study emphasizes the growing importance of technology in tourism. Tourists often leave reviews and ratings that reflect their experience. The authors used models like Support Vector Machines (SVM), Bidirectional LSTM, and Naïve Bayes to classify reviews as positive, negative, or neutral and predict ratings on a five-point scale. The analysis used data from a single source, limiting the generalizability of the results [1]. This paper examined how user-generated content (UGC), especially regarding hotel robots, can influence guest ratings. Naïve Bayes was used for prediction, and tools like Web Harvy and Python were applied to clean and process the data. Some irrelevant reviews may have been included, and there was limited diversity in robot-related data, which could affect model accuracy [2]. This study focused on analyzing customer reviews using deep learning techniques such as LSTM and GRUs. The models achieved high accuracy in predicting business outcomes based on guest feedback. Other models like Decision Tree, SVM, and Random Forest were also tested. The dataset was not well described, and conclusions lacked clear interpretation of results [3]. This article reviewed how traditional hotel star ratings compare with customer reviews. The authors discussed how voluntary star ratings are becoming more common and how they relate to online user feedback. The study focused on a small group of industry experts and did not include long-term data or insights into mandatory rating systems [4]. This article reviewed how traditional hotel star ratings compare with customer reviews. The authors discussed how voluntary star ratings are becoming more common and how they relate to online user feedback. The study focused on a small group of industry experts and did not include long-term data or insights into mandatory rating systems [5].

Methodology

Data Collection and Sample

Data was collected from a startup offering homestay services. The dataset related to customer information, stay details, and ratings. Fields include booking channel, room number, stay month, guest count, price paid, and ratings on cleanliness, accuracy, communication, and value for money.

Preprocessing Techniques

- Null values were handled using mean imputation for numerical fields and 'not available' placeholders for categorical values.
- Label Encoding was used to convert categorical fields into numeric values.
- Outliers were identified through boxplots and managed accordingly.

Exploratory Data Analysis

Graphs including histograms, pie charts, and violin plots were used to explore relationships. Key insights included seasonality in bookings and variance in ratings by platform.

Model Construction

Decision Tree Classifier was applied for classification tasks:

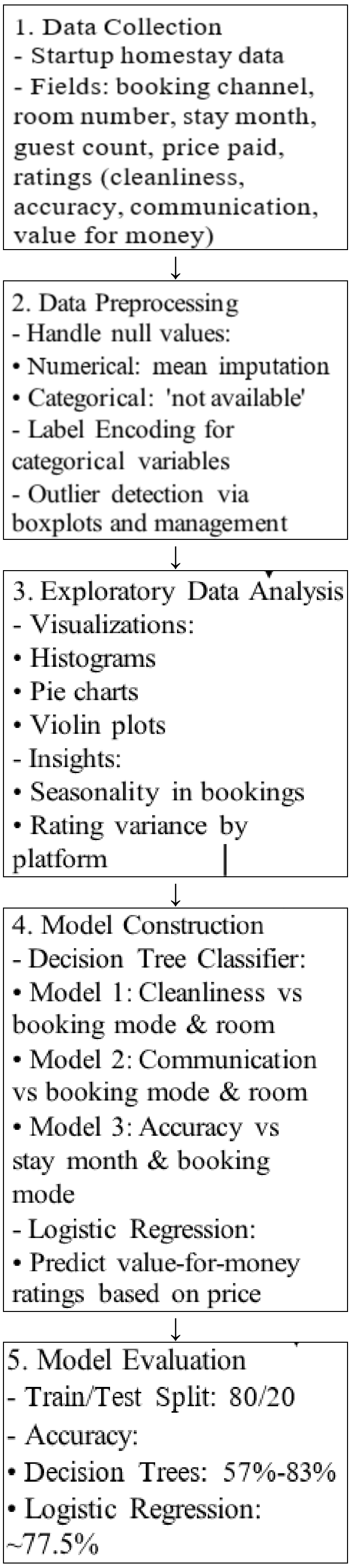
- **Model 1:** Cleanliness rating vs booking mode & room.
- **Model 2:** Communication rating vs booking mode & room.
- **Model 3:** Accuracy rating vs stays month & booking mode.

Logistic Regression predicted value-for-money ratings based on price.

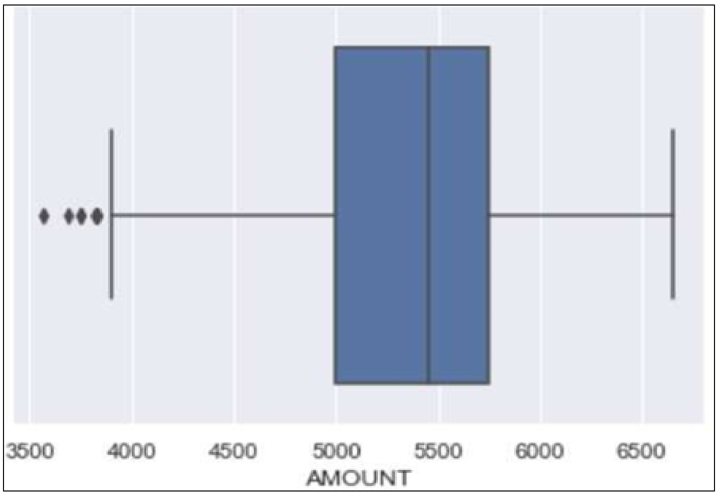
Evaluation Metrics

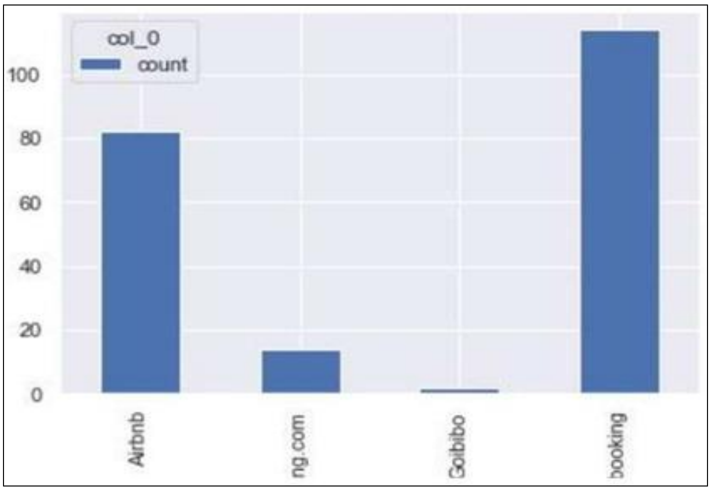
- **Train/test split:** 80/20.
- **Accuracy for Decision Tree models:** 57% to 83%.
- **Logistic regression accuracy:** ~77.5%.

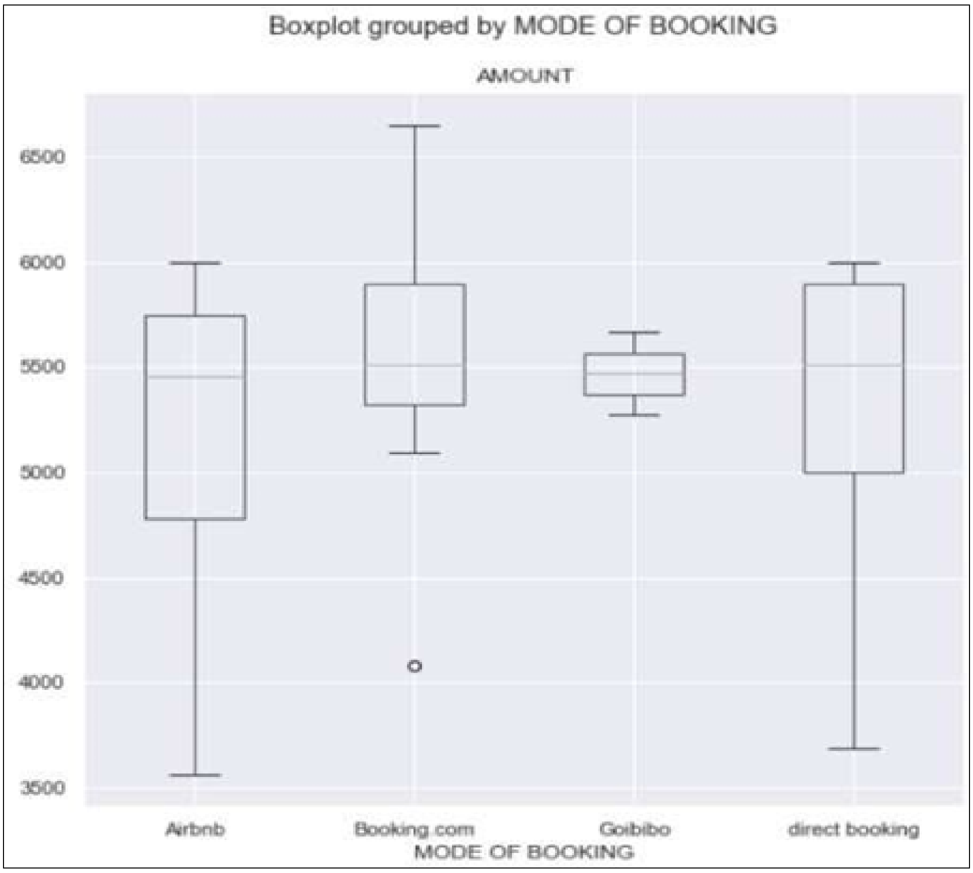
Flowchart

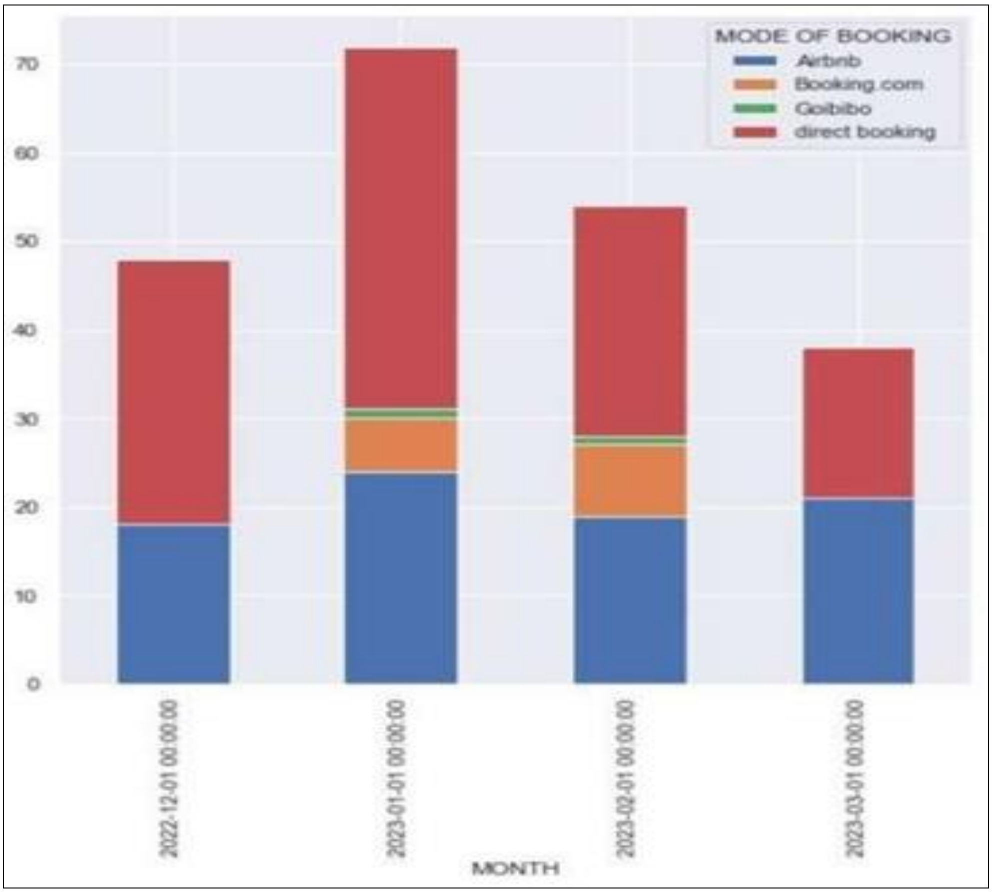


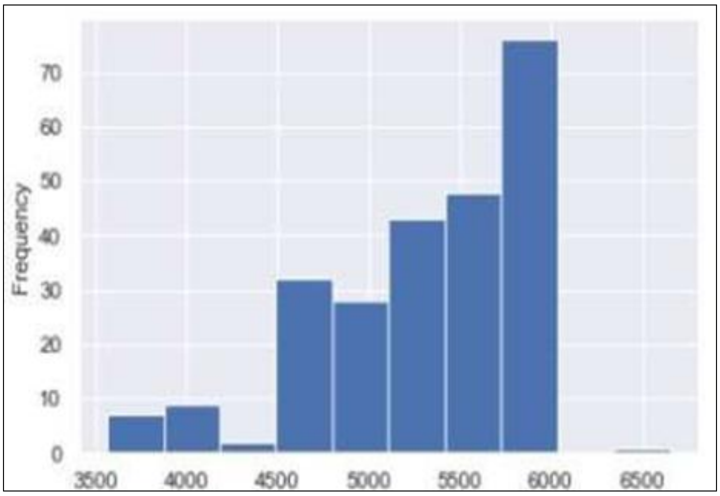
Results and Discussion

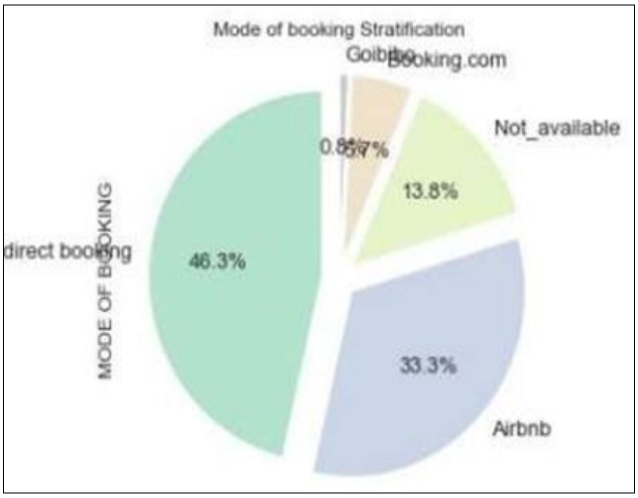


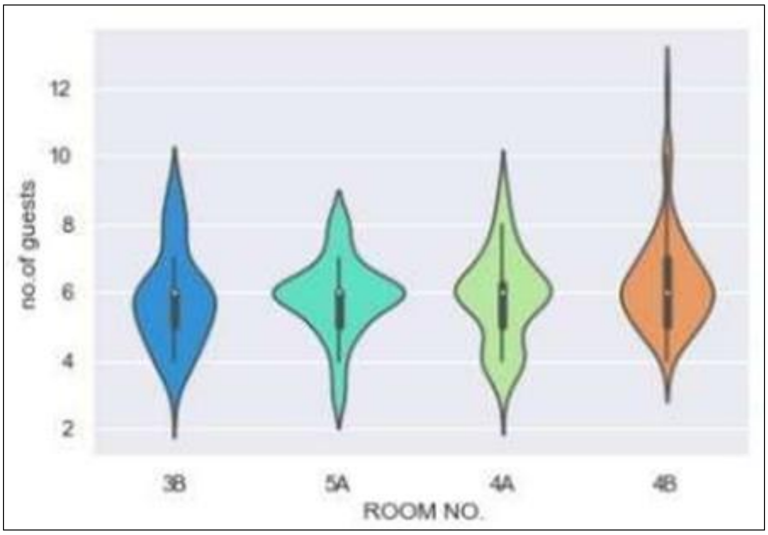


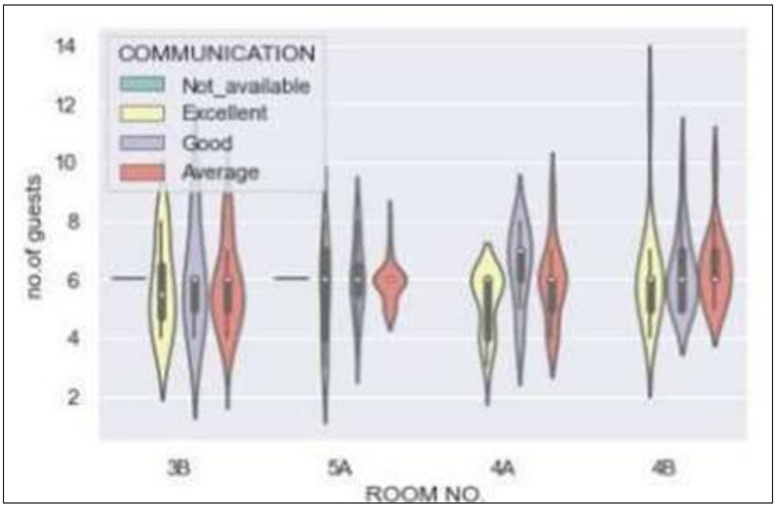



MODE OF BOOKING
Airbnb
Booking.com
Goibibo
direct booking
70
60
50
40
30
20
10
0
2022-12-01 00:00:00
2023-01-01 00:00:00
2023-02-01 00:00:00
2023-03-01 00:00:00
MONTH


Frequency
70
60
50
40
30
20
10
0
3500
4000
4500
5000
5500
6000
6500


Mode of booking Stratification
Goibibo
Booking.com
Not_available
direct booking
MODE OF BOOKING
46.3%
13.8%
33.3%
Airbnb


no.of guests
12
10
8
6
4
2
3B
5A
4A
4B
ROOM NO.



Decision Tree Model -1
Prediction of cleanliness rating based on mode of booking and Room number allotted.

```
In [119]:
dtmodel2.score(inputs2_n,targety2)
Out[119]:
0.573170731707317
```

Decision Tree Model 2
Prediction of Communication Ratings based on the Mode of booking and room number.

```
In [180]:
model.coef_
Out[180]:
array([[-0.00020093]])
In [181]:
model.intercept_
Out[181]:
array([-3.25739282e-08])
```

```
In [182]:
import math
def sigmoid(x):
  return 1 / (1 + math.exp(-x))
In [183]:
def prediction_function(AMOUNT):
    z = -0.00 * AMOUNT - 3.53
    y = sigmoid(z)
    return y
```

```
In [183]:
def prediction_function(AMOUNT):
    z = -0.00 * AMOUNT - 3.53
    y = sigmoid(z)
    return y
In [184]:
AMOUNT = 4700
prediction_function(AMOUNT)
Out[184]:
0.028470587690097256
```

Logistic Regression Model
Prediction of Value for money ratings with respect to the Amount/ price paid.
Each model highlighted different insights:

- Cleanliness was best predicted by room and booking mode (Accuracy: 82.9%).
- Communication ratings showed moderate predictability (57.3%), suggesting need for further features.
- Seasonal trends were observed in accuracy ratings (81.3%).
- The amount paid correlated well with perceived value (77.5%), affirming the relevance of pricing strategies.

These models provide a practical toolkit for hospitality providers to adjust their operations based on guest preferences and behavior.

Future Enhancements
The next stage of this research is the development of a real-time recommendation system for guests. This system would integrate:

- **Collaborative Filtering:** Recommending based on similar user behavior.
- **Content-Based Filtering:** Matching homestays with user preferences.
- **Sentiment Analysis:** using NLP on guest reviews to extract satisfaction indicators.
- **Matrix Factorization (SVD, ALS):** To uncover latent features influencing decisions.

These techniques will allow continuous model refinement as new data is generated, providing personalized, scalable, and intelligent guest services.

Conclusion
This paper has established a robust machine learning- based framework to predict service ratings in the hospitality sector. The models created demonstrate practical utility and offer significant value in enhancing guest satisfaction and operational decision-making. By integrating predictive analytics with future AI-driven recommendation systems, the study outlines a path toward intelligent, personalized, and data-driven hospitality services.

Limitations include a relatively small dataset and a lack of deep optimization for model hyperparameters. Future work will address these limitations with larger datasets, advanced modeling techniques, and live deployment scenarios.

References
1. Puh K, Babac MB. Predicting sentiment and rating of tourist reviews using machine learning. 2022.
2. Leal F, *et al*. A prediction and analysis of hotel ratings from crowd-sourced data. 2017.
3. Calero-Sanz J, *et al*. Predicting a new hotel rating system by analysing UGC content from TripAdvisor. 2022.
4. Hossen MS, *et al*. Hotel review analysis using deep learning. 2021.
5. Koutoulas D, *et al*. The present and future of hotel star ratings through the eyes of star rating operation. 2023.