SJIF Impact Factor (2026): 8.863| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 11 | Issue: 4 | April 2026 - Peer Reviewed Journal

AN EXPLAINABLE MACHINE LEARNING FRAMEWORK FOR QUALITY RATING PREDICTION OF VACATION RENTALS ON ONLINE TRAVEL PLATFORMS

Kumararaja Jetti[1], Gangiboina Yatheeswar[2], Grandhe Satwik[2], A.N.S.L. Sarvani[2], Karedla Venkata Sai Dhiraj Siddu[2], Mallela Vardhan Babu[2]

[1]Assistant Professor, Dept. of CSE (Cyber Security), BEC
[2]Students, Dept. of CSE (Cyber Security),BEC

Article DOI: https://doi.org/10.36713/epra26588
DOI No: 10.36713/epra26588

ABSTRACT

Online travel platforms offer a wide range of vacation rentals, making it difficult for users to choose the best option based on quality and reliability. This work presents a machine learning framework that predicts the quality rating of vacation rentals using factors such as location, price, amenities, host information, and customer reviews. The model analyses past data to identify patterns and generate accurate rating predictions. In addition to prediction, the framework explains how different features influence the final rating, helping users understand the reasons behind each score. This (Random Forest with SHAP (Shapley Additive Explanations).) approach supports travelers in making better booking decisions and helps property owners improve their services, leading to enhanced user satisfaction on online travel platforms.

KEYWORDS: *Vacation Rentals, Quality Rating Prediction, Machine Learning, Explainable Models, Customer Reviews, Feature Analysis, Online Travel Platforms, Data Analytics, User Satisfaction.*

I. INTRODUCTION

Online travel platforms have changed the way people plan their trips and book accommodation. Instead of relying on travel agents or limited hotel options, travelers can now explore a wide range of vacation rentals such as apartments, villas, and homestays. These platforms provide detailed information, including photos, descriptions, prices, and customer reviews. While this makes booking more convenient, it also creates confusion because users must choose from many options with different quality levels. One of the main factors that influence a traveler's decision is the quality rating of a property. Ratings are usually given by previous customers based on their experience. However, these ratings can sometimes be inconsistent or influenced by personal opinions, which makes it difficult for new users to fully trust them. A property with a high rating may still have hidden issues, while a lower-rated property might actually offer good value. This creates a need for a more reliable way to predict and understand property quality.

Machine learning provides a useful way to solve this problem by analyzing past data and identifying patterns that affect ratings. By studying factors such as location, price, amenities, host behavior, and customer feedback, a system can learn how these features contribute to the final rating. This allows the prediction of quality ratings for new or less-reviewed properties, helping users make better decisions based on data rather than guesswork. In addition to prediction, it is also important to understand why a particular rating is given. Many systems only provide results without explaining the reasons behind them, which can reduce user trust. Therefore, this work focuses on building a framework that not only predicts ratings but also explains how different factors influence the outcome. This makes the system more transparent and easier for users to understand.

The proposed framework can benefit both travelers and property owners. Travelers can choose better accommodations with more confidence, while property owners can learn which aspects of their service need improvement. By providing both prediction and clear explanations, this approach aims to improve the overall experience on online travel platforms and support better decision-making for all users.

II. LITERATURE SURVEY

The study titled *"Mining the Stars: Learning Quality Ratings with User-Facing Explanations for Vacation Rentals"* by A. Kornilova and L. Bernardi (2021) focuses on predicting the quality ratings of vacation rentals and clearly explaining those predictions to users. The main goal of this work is to build a system that can automatically estimate how a property would be rated based on available data such as reviews and listing details. At the same time, the system provides simple explanations so that users can understand why a certain rating is given. The authors treat existing user ratings as the target values and use them to train the model. During this

SJIF Impact Factor (2026): 8.863| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 11 | Issue: 4 | April 2026 - Peer Reviewed Journal

process, they address several challenges, such as handling different types of data and making the explanations easy for users to follow. Overall, the study aims to make rating predictions more reliable and transparent, helping both travelers and property owners make better decisions [1].

The paper titled *"An Ensemble Machine Learning Framework for Airbnb Rental Price Modeling without Using Amenity-Driven Features"* by I. Ghosh, R. K. Jana, and M. Z. Abedin (2023) presents a strong and scalable approach to predict rental prices on Airbnb. The main aim of the study is to develop a reliable model that can estimate prices even without depending on detailed amenity information. The authors use a combination of multiple machine learning models to improve prediction accuracy and make the system more robust. Their approach also focuses on improving the overall service quality of the Airbnb platform by helping hosts set better prices and assisting users in making informed choices. In addition, the study tries to make the model more understandable by explaining how different factors influence price predictions, even though such models are usually complex. Overall, the work highlights a practical way to predict rental prices effectively while keeping the system useful and easy to understand [2].

The paper titled *"What Online Review Features Really Matter? An Explainable Deep Learning Approach for Hotel Demand Forecasting"* by D. Zhang and C. Wu (2023) focuses on understanding how online reviews influence hotel demand. The study uses both numerical ratings and the sentiments expressed in customer reviews to predict future demand for hotels. Data is collected from online travel platforms where users share their experiences and feedback. The authors develop a deep learning model that not only makes accurate predictions but also helps identify which features, such as positive or negative opinions, have the most impact on demand. By analyzing customer reviews in detail, the study provides useful insights into traveler preferences and behavior. This helps hotel managers and travel platforms better understand what customers value, allowing them to improve services and plan more effectively [3].

The study titled *"Predicting Airbnb Rental Prices Using Customer Reviews Beyond Structured Data"* by L. Qin focuses on improving price prediction by using customer reviews along with basic property information. Since Airbnb is an online platform, travelers often depend on reviews from previous guests to understand the quality of a rental. This work highlights the importance of using not only structured data like price, location, and number of rooms, but also unstructured data such as written reviews. By analyzing these reviews, the model can capture deeper insights about customer experiences. The approach also aims to make the prediction process more understandable by showing how review-based information affects the final price. Overall, the study demonstrates that including customer opinions can lead to better predictions and more meaningful insights for both users and property owners [4].

The paper titled *"DeepPredict: A Zone Preference Prediction System for Online Lodging Platforms"* by Y. Ma and others (2021) presents a system that predicts which locations or zones users are likely to prefer when booking accommodations. The main goal is to improve the travel experience by helping users find suitable areas based on their interests and past behavior. The study analyzes user data, including previous bookings and reviews, to understand individual preferences. It also examines the relationship between users and their reviews to capture patterns in choices. Using a deep learning approach, the system learns these patterns and predicts preferred zones for future bookings. This helps travelers make quicker and better decisions, while also supporting platforms in offering more personalized recommendations [5].

III. METHODOLOGY

The proposed method uses a combination of the Random Forest algorithm and SHAP (Shapley Additive Explanations) to predict and explain the quality rating of vacation rentals. The process starts with collecting data from online travel platforms, which includes details such as location, price, number of rooms, amenities, host information, and customer reviews. This data is then cleaned to remove missing or incorrect values, and the text reviews are converted into useful numerical form so that they can be used by the model. After preprocessing, important features that influence the rating are selected. These may include factors like cleanliness, location advantage, pricing, and user feedback. The prepared data is then divided into training and testing sets. The Random Forest algorithm is used to train the model, where multiple decision trees are created and combined to produce a more accurate and stable prediction. This helps in reducing errors and improving overall performance.

Once the model is trained, it is used to predict the quality rating of new or unseen vacation rental listings. However, instead of only providing the predicted value, the framework also focuses on explaining the result. For this purpose, SHAP is applied to the trained model. SHAP calculates the contribution of each feature to the final prediction, showing how different factors increase or decrease the rating.

The output of the system includes both the predicted rating and a clear explanation of the most important features affecting that rating. For example, it can show whether a high rating is mainly due to good location, positive reviews, or reasonable pricing. This makes the system more transparent and easier to understand for users. By combining accurate prediction with clear explanations, this approach supports travelers in making better booking decisions and helps property owners understand areas for improvement. As a result, it enhances overall user satisfaction and builds trust in online travel platforms.

SJIF Impact Factor (2026): 8.863| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 11 | Issue: 4 | April 2026 - Peer Reviewed Journal

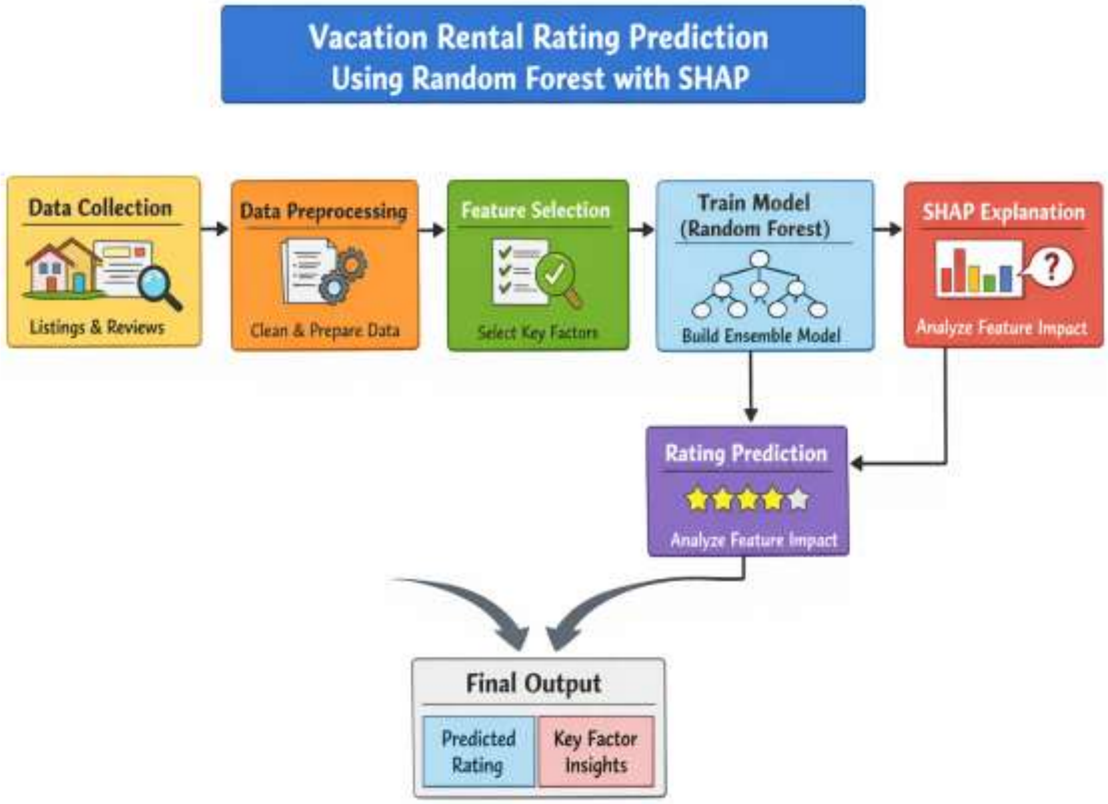

Figure 1: Overview of Proposed Work

The figure shows the complete flow of the proposed system for predicting vacation rental quality ratings. It starts with **data collection**, where information such as property details and customer reviews is gathered from online platforms. The next step is **data preprocessing**, where the collected data is cleaned and prepared for analysis. After that, **feature selection** is performed to identify the most important factors that influence the rating, such as price, location, and user feedback.

The selected features are then used to **train the model using the Random Forest algorithm**, which builds multiple decision trees to improve prediction accuracy. Once the model is trained, it generates the **rating prediction** for a given property. To make the result understandable, **SHAP explanation** is applied, which shows how each feature contributes to the final rating.

Finally, the system produces the **output**, which includes both the predicted rating and key factors influencing it. This helps users clearly understand the reason behind the rating and supports better decision-making.

IV. RESULTS

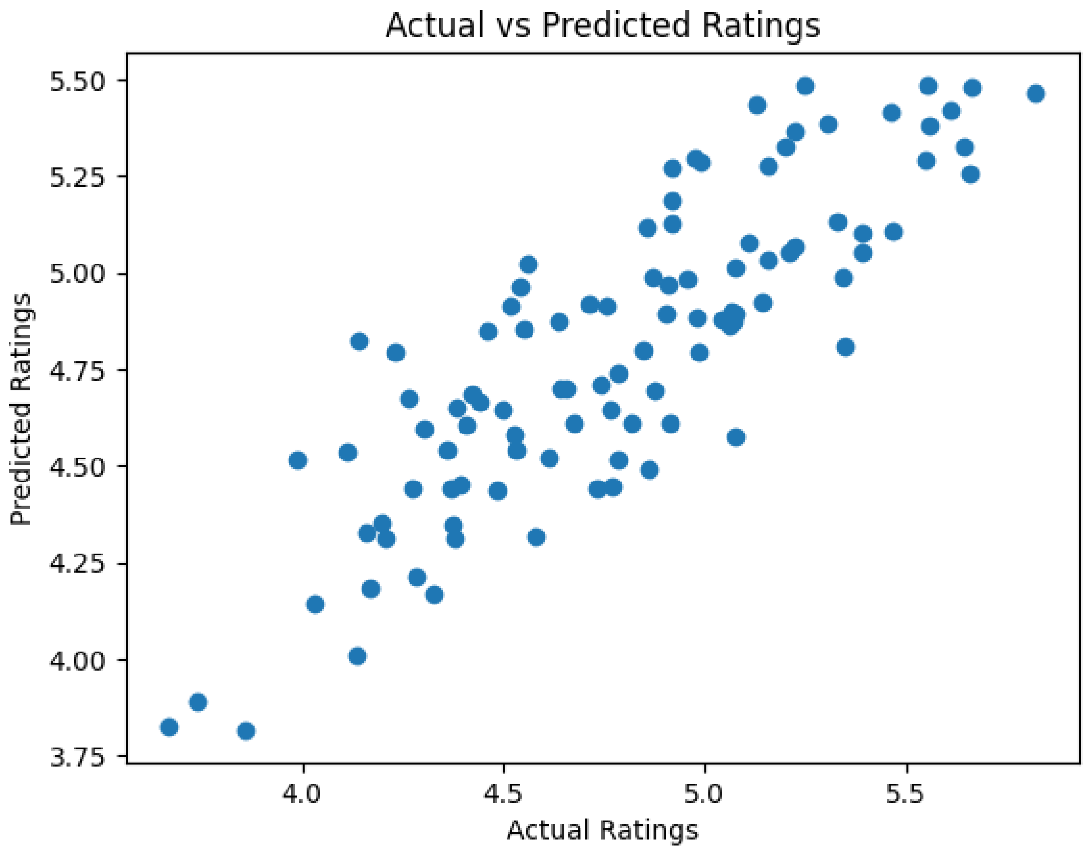

Figure 2: Actual vs Predicted Ratings

This figure shows the comparison between actual ratings and predicted ratings generated by the model. The points are closely aligned along a diagonal trend, indicating that the model predictions are close to the real values. This demonstrates that the model has good accuracy with only small prediction errors.

Figure 3: Error Distribution

This figure shows how the prediction errors are distributed in the model. The horizontal axis represents the difference between actual and predicted ratings (error), while the vertical axis shows how frequently those errors occur. Most of the errors are concentrated around zero, which means the model predictions are very close to the actual values in most cases. The spread of errors is small, indicating that there are no large mistakes. Overall, this figure shows that the model performs well with minimal prediction error.

SJIF Impact Factor (2026): 8.863| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 11 | Issue: 4 | April 2026 - Peer Reviewed Journal

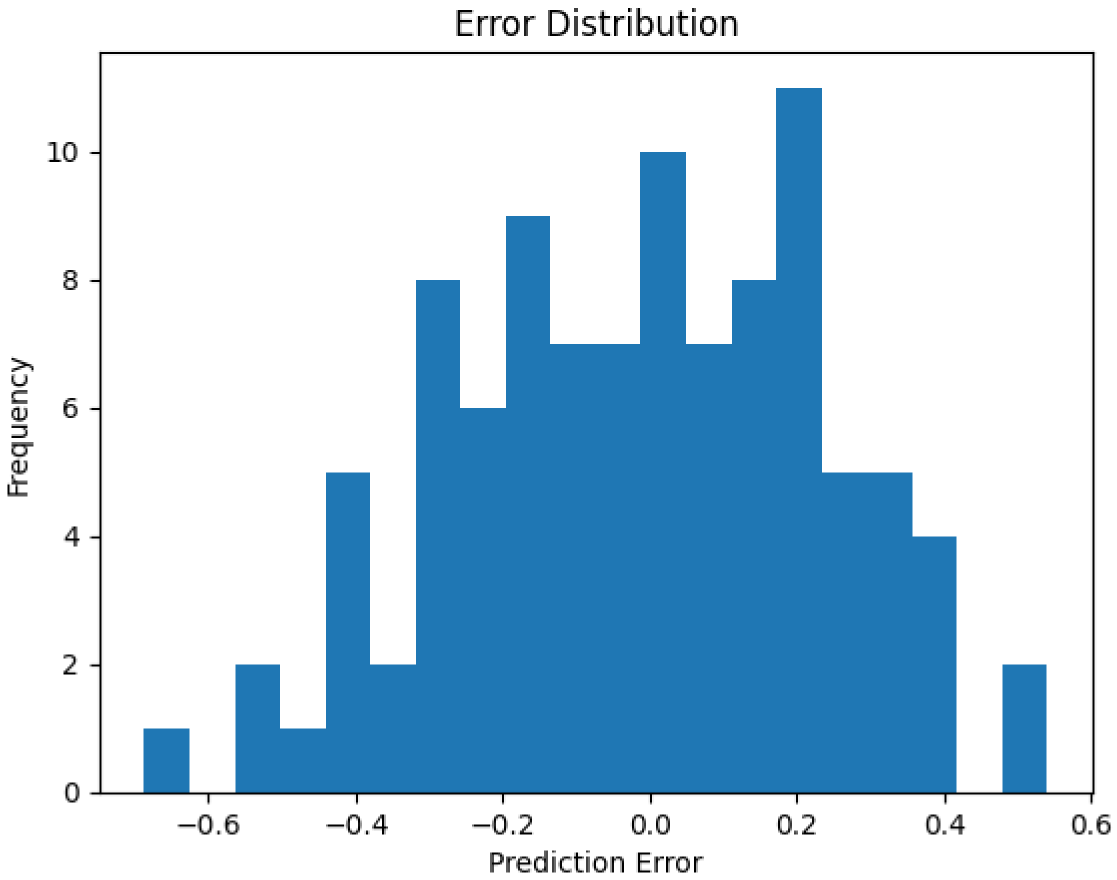

Figure 3: Error Distribution

This figure shows how the prediction errors are distributed in the model. The horizontal axis represents the difference between actual and predicted ratings (error), while the vertical axis shows how frequently those errors occur. Most of the errors are concentrated around zero, which means the model predictions are very close to the actual values in most cases. The spread of errors is small, indicating that there are no large mistakes. Overall, this figure shows that the model performs well with minimal prediction error.

SJIF Impact Factor (2026): 8.863| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 11 | Issue: 4 | April 2026 - Peer Reviewed Journal

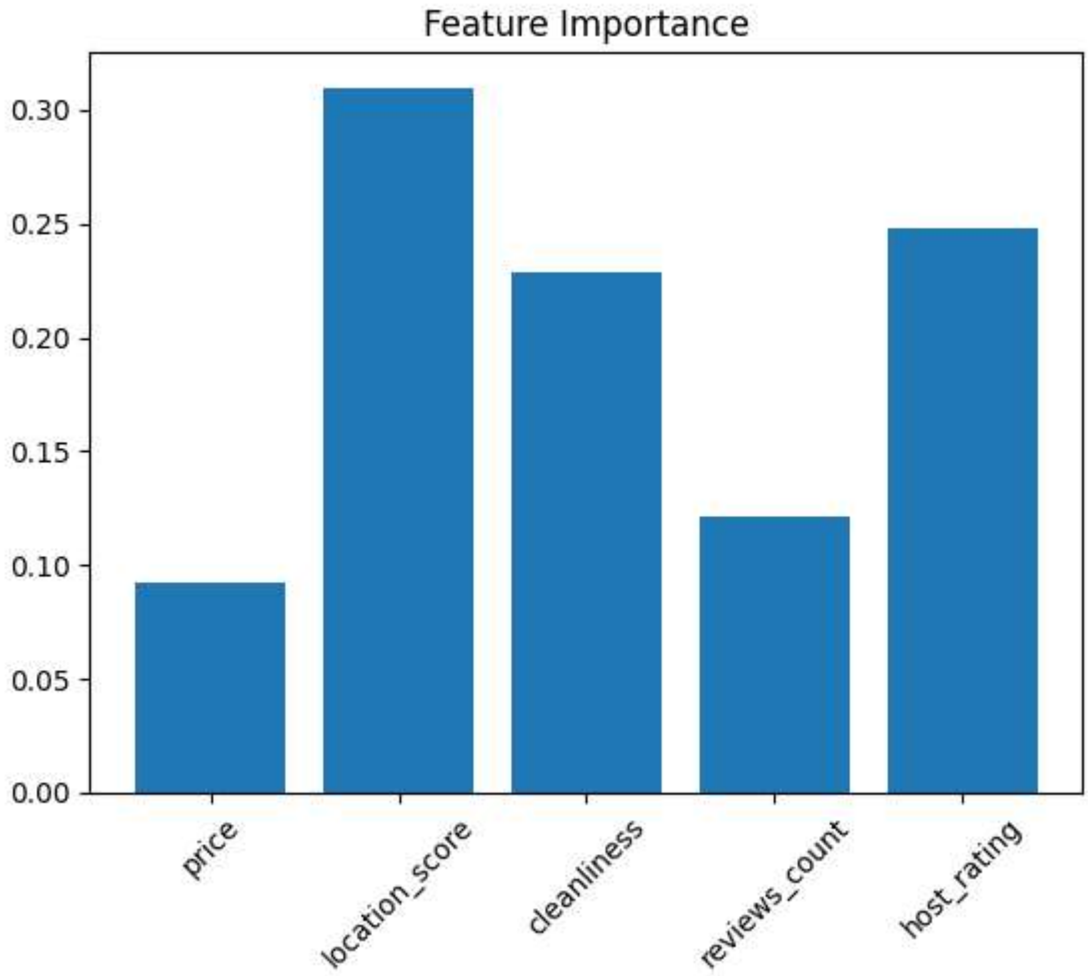

Figure 4: Feature Importance

This figure shows which factors have the most influence on predicting the rental rating. Each bar represents a feature, and its height indicates how important that feature is in the model's decision.

From the graph, **location score** has the highest importance, meaning it plays the biggest role in determining the rating. This is followed by **host rating** and **cleanliness**, which also strongly affect the prediction. On the other hand, **price** and **reviews count** have lower importance, indicating they have less impact compared to other factors.

SJIF Impact Factor (2026): 8.863| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 11 | Issue: 4 | April 2026 - Peer Reviewed Journal

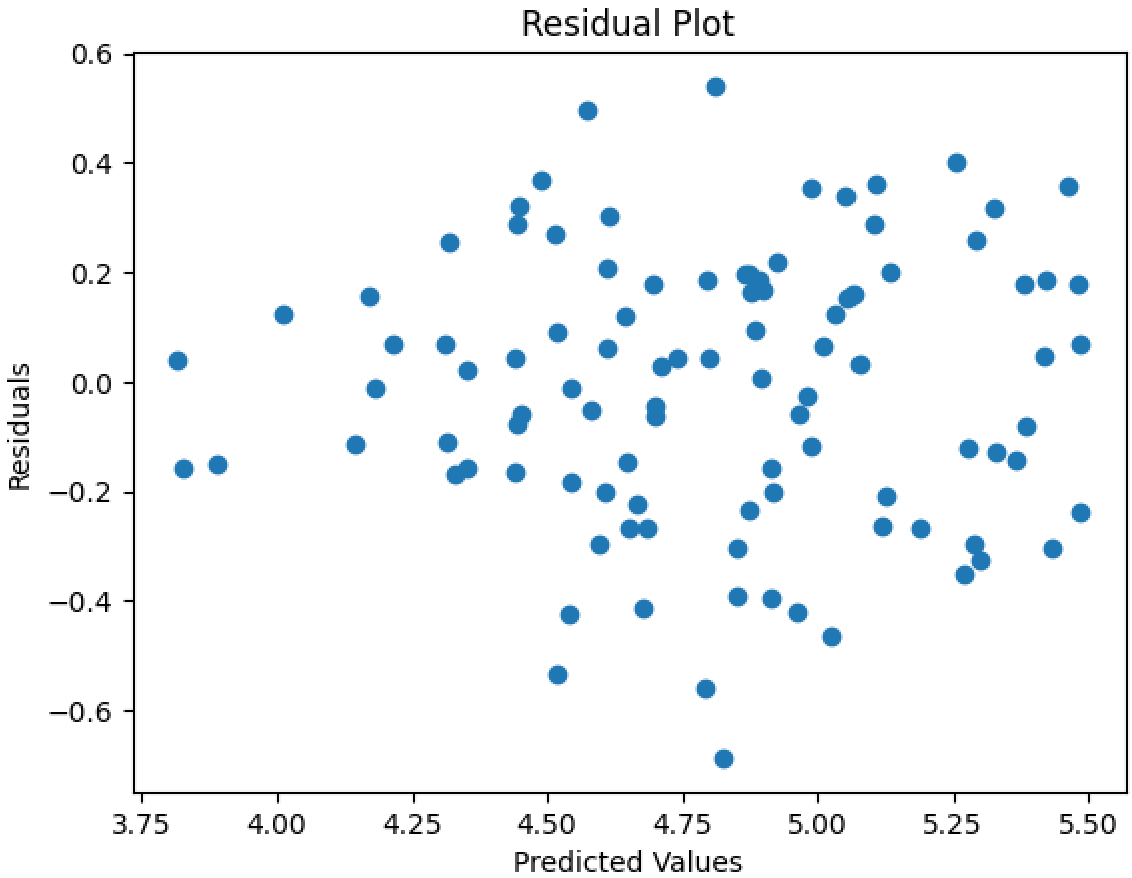

Figure 5: Residual Plot

This figure shows the residuals, which are the differences between the actual ratings and the predicted ratings, plotted against the predicted values. Each point represents one data sample.

The points are scattered randomly around the zero line, which indicates that the model does not have any strong bias in its predictions. There is no clear pattern or trend in the plot, meaning the model is performing consistently across different prediction values.

The spread of residuals is also relatively small, showing that the errors are not large. Overall, this figure confirms that the model fits the data well and makes reliable predictions without systematic mistakes.

This figure shows how each feature affects the model's prediction using SHAP values. Each row represents a feature, and each dot represents one data point. The horizontal axis shows the impact of that feature on the prediction, where values on the right increase the rating and values on the left decrease it.

The color of the dots indicates the value of the feature, where red means high values and blue means low values. From the plot, we can see that **location score**, **host rating**, and **cleanliness** have strong positive impacts when their values are high, meaning better scores in these features lead to higher predicted ratings. On the other hand, lower values of these features tend to reduce the rating. Features like **reviews count** and **price** have a smaller impact compared to others, but they still influence the prediction to some extent. Overall, this figure clearly explains how each feature contributes to the model's decision, making the prediction process more transparent and easy to understand.

SJIF Impact Factor (2026): 8.863| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 11 | Issue: 4 | April 2026 - Peer Reviewed Journal

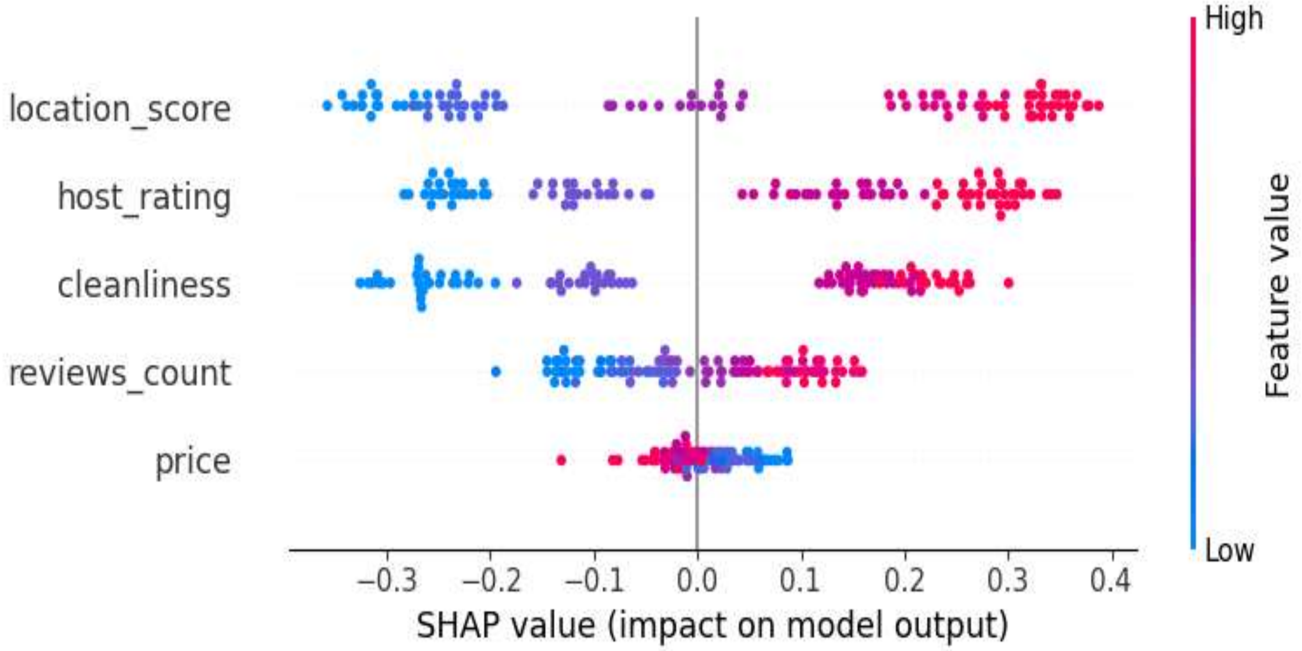

Figure 6: SHAP Summary Plot (Feature Impact on Prediction)

V. CONCLUSION

This work presented a machine learning framework to predict the quality ratings of vacation rentals using important features such as location, price, cleanliness, host rating, and customer reviews. The model was able to produce accurate predictions, as shown by the performance metrics and visual results. The comparison between actual and predicted values indicated that the model performs well with only small errors, and the error analysis confirmed that the predictions are consistent and reliable. In addition to prediction, the use of SHAP helped explain how each feature influences the final rating. It was observed that factors like location score, host rating, and cleanliness play a major role in determining the quality of a rental, while other factors have a smaller impact. This makes the model easy to understand and increases trust in its results.

REFERENCES

1. *Gao Jr, Y. (2025). Unravelling the What, How, and When of Airbnb Price Prediction: A Case Study of London Based on Machine Learning.*
2. *Tang, J., Cheng, J., & Zhang, M. (2024). Forecasting Airbnb prices through machine learning. Managerial and Decision Economics, 45(1), 148-160.*
3. *Kornilova, A., & Bernardi, L. (2021, March). Mining the stars: learning quality ratings with user-facing explanations for vacation rentals. In Proceedings of the 14th ACM International Conference on Web Search and Data Mining (pp. 976-983).*
4. *Ghosh, I., Jana, R. K., & Abedin, M. Z. (2023). An ensemble machine learning framework for Airbnb rental price modeling without using amenity-driven features. International Journal of Contemporary Hospitality Management, 35(10), 3592-3611.*
5. *Zhang, D., & Wu, C. (2023). What online review features really matter? An explainable deep learning approach for hotel demand forecasting. Journal of the Association for Information Science and Technology, 74(9), 1100-1117.*
6. *QIN, L. PREDICTING AIRBNB RENTAL PRICES USING CUSTOMER REVIEWS BEYOND STRUCTURED DATA (Doctoral dissertation, tilburg university).*
7. *Ma, Y., Sun, H., Chen, Y., Zhang, J., Xu, Y., Wang, X., & Hui, P. (2021). DeepPredict: A zone preference prediction system for online lodging platforms. Journal of Social Computing, 2(1), 52-70.*
8. *Gao Jr, Y. (2025). Unravelling the What, How, and When of Airbnb Price Prediction: A Case Study of London Based on Machine Learning.*