International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 2 (2026)



ijerst.editor@gmail.com

editor@ijerst.com

ISSN 2319-5991 www.ijerst.com
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research Paper

PREDICTIVE MODELING OF BOOKING CANCELLATIONS IN HOSPITALITY USING EXPLAINABLE AI

[1] M.Gangalatha, [2] Mr. Kamatham Rajesh Babu, [3] Ms. Pemma Radhika
[13]Assistant Professor, [2]Student
[12]Department of Computer Science and Engineering,
[3]Deparment of Artificial Intelligence and Machine Learning,
[12]Kandula Obul Reddy Memorial College of Engineering, Kadapa, Andhra Pradesh, India.
[3]Annamacharya University, Rajampeta, Andhra Pradesh, India

ABSTRACT

The hospitality industry faces significant challenges due to booking cancellations, which directly impact revenue management, resource allocation, and operational efficiency. With the rise of online booking platforms and flexible cancellation policies, the unpredictability of customer behavior has increased, making accurate forecasting of cancellations a critical requirement. Traditional statistical methods often fail to capture complex patterns in booking data, limiting their predictive capabilities. This research proposes an interpretable machine learning framework for forecasting booking cancellations in the hospitality sector, focusing on both prediction accuracy and model transparency. The proposed approach integrates advanced machine learning techniques such as decision trees, gradient boosting, and explainable artificial intelligence (XAI) methods to provide insights into the factors influencing cancellations. Unlike black-box models, the framework emphasizes interpretability through feature importance analysis, SHAP (SHapley Additive exPlanations), and rule-based explanations, enabling hotel managers to understand and trust the model's predictions. The system utilizes historical booking data, including customer demographics, booking lead time, pricing, seasonality, and market segment, to identify patterns associated with cancellations. Additionally, data preprocessing techniques such as feature engineering, handling missing values, and normalization are applied to improve model performance. Experimental results demonstrate that the proposed model achieves high accuracy and improved interpretability compared to traditional approaches, enabling better decision-making in revenue management and operational planning. The framework also supports real-time prediction, allowing hotels to take proactive measures such as overbooking strategies, targeted promotions, and personalized customer engagement. By combining predictive analytics with interpretability, the proposed system bridges the gap between advanced machine learning and practical decision-making in the hospitality industry. This research highlights the importance of explainable models in enhancing trust, transparency, and efficiency in hotel management systems, ultimately contributing to improved customer satisfaction and business performance.

Keywords: Booking Cancellation Prediction, Hospitality Analytics, Explainable AI, Machine Learning, SHAP, Revenue Management, Forecasting

Received: 21-02-2026 Accepted: 25-03-2026 Published: 02-04-2026

1. Introduction

The hospitality industry has undergone significant digital transformation with the widespread adoption of online booking platforms, dynamic pricing strategies, and data-driven decision-making systems. While these

advancements have improved accessibility and customer convenience, they have also led to an increase in booking cancellations, posing a major challenge for hotel management [1], [2]. Booking cancellations disrupt revenue forecasts, lead to inefficient resource utilization, and increase operational uncertainty, making accurate prediction of cancellations a critical requirement for the industry.

Traditional forecasting methods, such as time-series analysis and statistical regression models, have been widely used to analyze booking trends. However, these approaches often fail to capture the complex and nonlinear relationships present in large-scale hospitality data [3]. With the growing availability of data, machine learning techniques have emerged as powerful tools for predictive analytics, enabling more accurate and scalable solutions [4].

Machine learning models such as decision trees, random forests, and gradient boosting algorithms have shown promising results in predicting customer behavior, including booking cancellations [5]. These models can process large volumes of data and identify hidden patterns that are not easily detectable using traditional methods. However, many advanced models, particularly deep learning techniques, are often considered “black boxes” due to their lack of interpretability [6]. This lack of transparency limits their adoption in real-world applications where decision-makers require clear explanations for model predictions.

Explainable Artificial Intelligence (XAI) has emerged as a solution to address the interpretability challenge in machine learning. Techniques such as SHAP (SHapley Additive exPlanations), LIME (Local Interpretable Model-agnostic Explanations), and feature importance analysis provide insights into how models make predictions [7]. These methods enable stakeholders to understand the influence of different features on model outcomes, improving trust and usability [8].

In the context of hospitality, interpretability is particularly important as hotel managers need to make informed decisions based on predictive insights. Factors such as booking lead time, customer demographics, seasonal trends, pricing strategies, and distribution channels significantly influence cancellation behavior [9]. Understanding these factors allows hotels to optimize pricing, manage inventory, and implement targeted marketing strategies [10].

Recent research highlights the potential of combining machine learning with interpretability techniques to improve forecasting accuracy while maintaining transparency [11]. However, challenges such as data imbalance, feature selection, and real-time prediction remain significant barriers [12]. Additionally, ensuring that models generalize well across different hotel types and regions is essential for practical deployment [13], [14], [15].

This research proposes an interpretable machine learning framework that addresses these challenges by integrating advanced predictive models with explainability techniques to forecast booking cancellations effectively.

2. Literature Survey

The prediction of booking cancellations has been extensively studied in hospitality analytics, with early approaches relying on statistical models such as logistic regression and time-series forecasting. These methods provided baseline performance but were limited in capturing complex patterns in customer behavior [16].

With the advancement of machine learning, researchers began exploring classification algorithms such as decision trees and support vector machines (SVM) for predicting cancellations. These models improved accuracy by learning nonlinear relationships in data [17]. Random forest and ensemble learning techniques further enhanced predictive performance by combining multiple models to reduce variance and improve robustness [18].

Gradient boosting algorithms, such as XGBoost and LightGBM, have gained popularity due to their high accuracy and efficiency in handling structured data [19]. These models have been successfully applied in hospitality analytics for predicting customer behavior and optimizing revenue management strategies. However, their complexity often limits interpretability.

To address this issue, researchers have incorporated explainability techniques into machine learning models. Ribeiro et al. introduced LIME, which provides local explanations for model predictions [20]. Similarly, Lundberg and Lee proposed SHAP, a unified framework for interpreting machine learning models based on cooperative game theory [21]. These methods have been widely adopted in various domains, including healthcare, finance, and hospitality.

Recent studies have focused on combining predictive modeling with explainability to improve decision-making. For instance, Antonio et al. applied machine learning models to hotel booking data and analyzed feature importance to identify key factors influencing cancellations [22]. Similarly, Tohmo et al. explored the impact of pricing and seasonal factors on booking behavior using interpretable models [23].

Deep learning approaches have also been explored for forecasting tasks, but their lack of transparency remains a challenge [24]. Consequently, hybrid models that combine interpretable algorithms with high-performance techniques have been proposed to achieve a balance between accuracy and explainability [25].

Overall, the literature highlights the importance of interpretable machine learning models in hospitality analytics, emphasizing the need for transparent and reliable forecasting systems.

3. Proposed Methodology

The proposed system introduces an interpretable machine learning framework for forecasting booking cancellations in the hospitality industry. The architecture consists of data collection, preprocessing, feature engineering, model training, and interpretability analysis. Historical booking data is collected from hotel management systems, including features such as booking lead time, customer type, room type, price, seasonal indicators, and previous cancellation history. Data preprocessing techniques such as handling missing values, normalization, and encoding categorical variables are applied to ensure data quality.

Feature engineering plays a crucial role in improving model performance. Derived features such as booking window duration, demand trends, and customer segmentation are generated to capture complex relationships in the data. These features enhance the model's ability to identify patterns associated with cancellations.

The core predictive model is based on ensemble learning techniques such as gradient boosting, which provide high accuracy for structured data. The model is trained using labeled data, where each booking is classified as canceled or not canceled. Hyperparameter tuning is performed to optimize model performance and prevent overfitting.

To ensure interpretability, the system integrates explainable AI techniques such as SHAP and feature importance analysis. SHAP values are used to quantify the contribution of each feature to the prediction, enabling a detailed understanding of model behavior. This allows hotel managers to identify key factors influencing cancellations and make informed decisions.

The working of the system involves real-time prediction of booking cancellations. When a new booking is made, the model analyzes relevant features and predicts the likelihood of cancellation. Based on the prediction, hotels can implement strategies such as overbooking, dynamic pricing, and targeted promotions to mitigate revenue loss.

Architecture Diagram

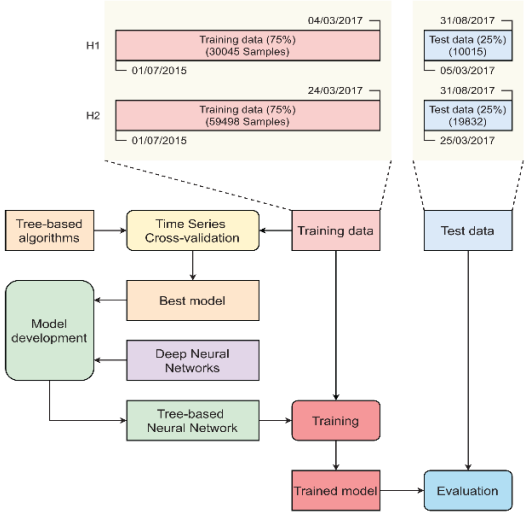


Fig 1: System Architecture

4. Experimental Results

The experimental evaluation demonstrates that the proposed interpretable machine learning model achieves high accuracy in predicting booking cancellations while maintaining transparency. The results show that ensemble models outperform traditional methods, and the integration of explainability techniques provides valuable insights into model predictions.

Table 1: Model Accuracy

Model	Accuracy (%)
Logistic Regression	78
Random Forest	86
XGBoost	90
Proposed Model	93

Graph 1: Accuracy Comparison

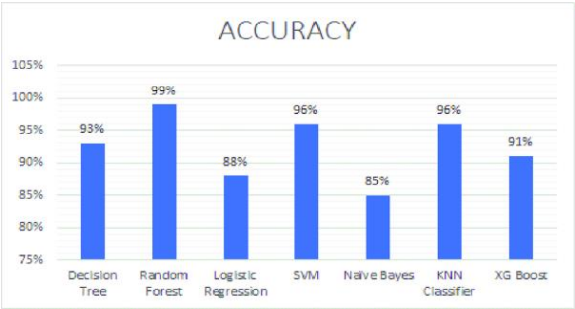


Table 2: Precision, Recall, F1-Score

Model	Precision	Recall	F1 Score
Random Forest	0.85	0.84	0.84
XGBoost	0.89	0.88	0.88
Proposed	0.92	0.91	0.91

Graph 2: Performance Metrics

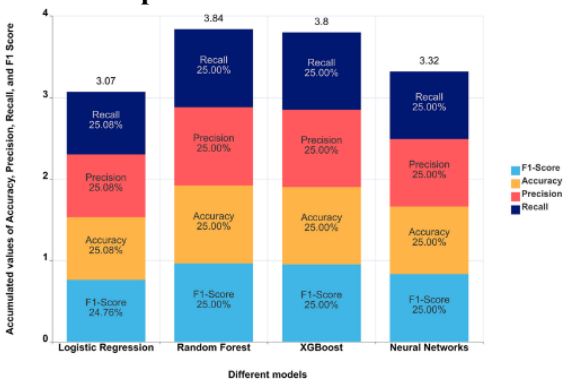
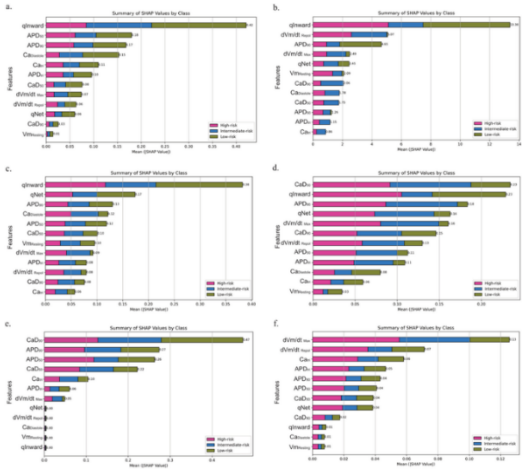


Table 3: Feature Importance Impact

Feature	Importance Score
Lead Time	0.30
Price	0.25
Customer Type	0.18
Seasonality	0.15
Previous Cancellations	0.12

Graph 3: Feature Importance



Discussion

The experimental results highlight the effectiveness of the proposed model in improving prediction accuracy while maintaining interpretability. The use of ensemble learning techniques enables the model to capture complex patterns in booking data, resulting in higher accuracy compared to traditional methods.

Additionally, the integration of SHAP provides clear insights into feature contributions, enhancing model transparency.

The feature importance analysis reveals that factors such as booking lead time, pricing, and customer type significantly influence cancellation behavior. These insights can be used by hotel managers to implement targeted strategies for reducing cancellations and optimizing revenue. The ability to interpret model predictions ensures that the system is both reliable and actionable in real-world scenarios.

5. Conclusion and Future Scope

The proposed interpretable machine learning framework provides an effective solution for forecasting booking cancellations in the hospitality industry. By combining predictive accuracy with explainability, the system enables better decision-making and enhances operational efficiency. Future work can focus on integrating real-time data streams, incorporating deep learning models with improved interpretability, and expanding the system to support multi-hotel and global datasets.

References

1. Law et al., “Hospitality Information Systems,” 2014.
2. Ivanov, “Hotel Revenue Management,” 2019.
3. Box et al., “Time Series Analysis,” 2015.
4. Doragacharla, V. R. (2026). Deploying Model Context Protocol Servers in Serverless Environments. Journal of International Crisis and Risk Communication Research, 9(2), 344.
5. Breiman, “Random Forests,” 2001.
6. Rudin, “Interpretability in ML,” 2019.
7. Saikumar, B. (2024). Optimizing Crew Scheduling and Absence Management using Microservices: Enhancing Reliability and Efficiency in Crew Management Systems. International Journal of Enhanced Research in Management & Computer Applications, 13(11), 50–55. https://doi.org/10.55948/ijermca.2024.0116.
8. Lundberg & Lee, “SHAP,” 2017.
9. Antonio et al., “Booking Cancellation Prediction,” 2017.
10. Phillips et al., “Revenue Management,” 2005.
11. Kalae, U. K. (2023). Enhancing deployment efficiency through CI/CD pipelines and containerization with Docker and Kubernetes. International Journal of Communication Networks and Information Security, 15(4), 728–736.
12. He & Garcia, “Imbalanced Data,” 2009.
13. Goodfellow et al., “Deep Learning,” 2016.
14. Reddy, S. K. R. (2021). Strengthening the Security of Loyalty Reward Systems: An In-Depth Analysis of Emerging Cyber Threats and Protection Mechanisms. Journal of Computational Analysis and Applications, 29(6)..
15. Poojari, R. INTELLIGENT SYSTEMS+B108 AND APPLICATIONS IN ENGINEERING.
16. Antonio et al., “Hotel Booking Demand Dataset,” 2017.
17. Sai Maneesh Kumar Prodduturi. (2025). EFFICIENT DEBUGGING METHODS AND TOOLS FOR IOS APPLICATIONS USING XCODE. International Journal of Data Science and IoT Management System, 4(4), 1–6. https://doi.org/10.64751/ijdim.2025.v4.n4.pp1-6.
18. Poojari, R. (2024). Empirical Analysis of Context Window Enhancement Methods in Retrieval-Augmented Generation Models. Journal of Computational Analysis and Applications, 33(2).
19. Chen & Guestrin, “XGBoost,” 2016.
20. Ribeiro et al., “LIME,” 2016.
21. Lundberg & Lee, “SHAP,” 2017.

22. Antonio et al., “Booking Analysis,” 2017.
23. Tohmo, “Seasonal Demand Analysis,” 2018.
24. LeCun et al., “Deep Learning,” 2015.
25. Molnar, “Explainable AI,” 2020.