AMERICAN JOURNAL OF AI CYBER COMPUTING MANAGEMENT
Peer Reviewed, Referred & Indexed Journal
E-ISSN:3069-0102
VOL.6, NO. 2(2026)

Machine Learning in Hospitality: Interpretable Forecasting of Booking Cancellations

Ms.N.Pushpanjali, MCA -Assistant Professor, Department of MCA, Bapatla Engineering College Bapatla, Andhra Pradesh

Ms.Manam Vani, (Reg No : Y25MC23045), Mr.Aritakula Naga Venkata Sai Krishna, (Reg No: Y25MC23004)
Mr.Kattupalli Vinod, (Reg No: Y25MC23033), Mr.Malapati Venkateswarlu, (Reg No: Y25MC23043)
Department of MCA, Bapatla Engineering College, Bapatla, Andhra Pradesh, India

***Abstract*—The hospitality industry has a lot of problems with money and how things run because people cancel their bookings at the minute. This affects how money the hotels make and how they plan their services. Now that we have a lot of booking information we can use special computer programs to predict when people will cancel.**

This study is about making a computer program that can predict when people will cancel their hotel bookings. We use information about the bookings and the people who made them to make predictions. We try out computer programs like Logistic Regression and Decision Trees to see what makes people cancel.

We want to know why people cancel so we look at things like how it takes for people to cancel what kind of customers they are, if they paid a deposit and if they have booked with us before. This helps hotel managers understand when people might cancel and make plans to stop it from happening. They can change their prices control how bookings they take and talk to the customers who are most likely to cancel.

We test our computer program. It works well and is easy to understand. This means that hotel managers can use it to make decisions. Our results show that using these kinds of computer programs can really help the hospitality industry make choices based on data.

The hospitality industry can use this to make things better, for themselves and their customers. Hotel booking cancellations are a problem but with the right tools we can predict when they will happen and do something about it. Hotel managers can use this information to make their hotels better and make money. The computer program is helpful because it can predict hotel booking cancellations and help hotel managers understand why people cancel.

Keywords: Machine Learning, Booking Cancellation Prediction, Hospitality Analytics, Interpretable Models, Predictive Analytics, Hotel Reservation Systems.

I. INTRODUCTION

The hotel industry has changed a lot with the use of booking platforms and reservation systems. Hotels now use platforms to manage bookings, prices and customer relationships. Online booking systems are convenient for customers. They have also led to more booking cancellations. High cancellation rates cause problems for hotels, such as lost revenue, wasted resources and difficulty predicting demand.

Predicting booking cancellations is now a task for hotel managers. Traditional forecasting methods do not work well with booking data. With more reservation data available machine learning techniques are becoming useful for finding patterns and predicting customer behavior. Machine learning models look at factors like booking lead time, customer type, market segment, pricing policies and previous booking history to estimate the chance of cancellation.

However some advanced machine learning models are hard to understand making it difficult for hotel managers to trust the predictions. In real-world decision-making it is essential to understand the factors influencing predictions. Interpretable machine learning techniques help show the important features contributing to booking cancellations enabling managers to make informed decisions.

This study focuses on creating a machine learning framework for predicting hotel booking cancellations. The approach uses supervised learning algorithms to analyze reservation data and predict cancellation behavior while keeping the model transparent. By identifying factors that influence cancellations the model can help hotel managers implement proactive strategies such as dynamic pricing, improved booking policies and optimized overbooking management.

The findings of this research show how interpretable machine learning can support data-driven decision making in the hospitality sector. By combining accuracy, with model transparency the approach provides practical insights that can help hospitality businesses reduce uncertainty and improve operational efficiency.

II. LITERATURE SURVEY

Forecasting demand and booking behavior has been an important research topic in the hospitality and tourism industry. Several studies have explored statistical and machine learning techniques to improve prediction accuracy and

…

Received: 28-02-2026 | Accepted: 02-04-2026 | Published: 19-04-2026 | www.ajaccm.com

support hotel management decisions.

Chow et al. [1] presented one of the early forecasting systems for hotel occupancy rates using integrated ARIMA models. Their study focused on capturing seasonal and trend components present in historical occupancy data. The proposed model showed better forecasting performance compared to traditional statistical techniques. However, the model relies on linear assumptions and cannot effectively capture complex customer behavior patterns. Furthermore, the study does not address booking cancellation prediction or model interpretability.

Chu [2] investigated the use of ARMA-based time-series models for forecasting tourism demand. The research analyzed different ARMA configurations to model variations in tourist arrivals. The results indicated that time-series statistical models can provide reliable short-term forecasts when adequate historical data is available. Nevertheless, these models struggle to handle non-linear patterns and sudden behavioral changes in demand. Additionally, the study does not incorporate machine learning techniques or interpretability mechanisms.

Frechtling [3] provided a comprehensive overview of tourism demand forecasting techniques, including econometric models, time-series methods, and judgmental forecasting approaches. The work highlights the importance of accurate forecasting for effective strategic planning within the tourism and hospitality sectors. Although the book establishes a strong theoretical foundation for forecasting, it mainly focuses on traditional statistical methods. Modern data-driven approaches such as machine learning and explainable models are not extensively discussed, and booking cancellation behavior is not specifically analyzed.

Kourentzes et al. [4] examined the role of temporal aggregation in improving demand forecasting accuracy. Their study demonstrated that using multiple aggregation levels can enhance forecasting performance, particularly when dealing with noisy demand data. While the approach provides valuable insights into demand forecasting methodologies, it is not specifically designed for the hospitality domain or booking cancellation prediction. Moreover, the framework does not include machine learning models or explainable techniques for decision support.

Li et al. [5] proposed a hybrid forecasting model that combines Principal Component Analysis (PCA) with an improved Backpropagation Neural Network (BPNN) to predict tourist volume. By integrating online search data from the Baidu Index, the model improved forecasting accuracy and captured complex non-linear patterns. Despite its strong predictive performance, the neural network operates as a black-box model, making it difficult to interpret the factors influencing predictions. In addition, the study focuses on tourist demand forecasting rather than booking cancellation behavior.

Antonio et al. [6] conducted one of the earliest studies applying machine learning techniques to predict hotel booking cancellations. The authors utilized classification algorithms to determine whether a booking would be canceled based on historical reservation data. Their results demonstrated that machine learning models outperform traditional statistical approaches in predicting cancellations. However, the research primarily emphasizes prediction accuracy and does not explore model interpretability or provide insights into the behavioral factors behind cancellations.

Sa´nchez-Medina and Sa´nchez [7] further explored the use of machine learning and big data techniques for forecasting hotel booking cancellations. Their study demonstrated the effectiveness of large-scale data analytics in improving prediction performance. The research highlighted the operational benefits of cancellation forecasting for hotel revenue management. Nevertheless, the models used in the study lack transparency and explainability, limiting their usefulness for decision-making processes that require interpretable insights.

Overall, previous studies have made significant contributions to forecasting demand and predicting booking cancellations in the hospitality industry. However, many existing approaches either rely on traditional statistical models that cannot capture complex patterns or utilize machine learning models that lack interpretability. Therefore, there is a need for an interpretable machine learning framework that can accurately forecast booking cancellations while providing clear insights into the factors influencing customer behavior.

III. ALGORITHM

Input: Booking dataset $D = \{(x_i, y_i)\}_{i=1}^{N}$ where x_i represents booking features and $y_i \in \{0, 1\}$ indicates cancellation status.

Output: Trained prediction model M and feature importance explanations E.

Step 1: Data Collection

Collect historical hotel booking data including attributes such as lead time, customer type, deposit type, booking channel, previous cancellations, and special requests.

Step 2: Data Preprocessing

1) Remove missing and inconsistent records.
2) Encode categorical variables using one-hot encoding.
3) Normalize numerical features.

Let the processed dataset be:

$$D' = \{(x_i', y_i)\}_{i=1}^{N} \quad (1)$$

Step 3: Dataset Partitioning

Received: 28-02-2026 | Accepted: 02-04-2026 | Published: 19-04-2026 | www.ajaccm.com

Split D' into training and testing sets:

$$D_{train} = 80\%, \quad D_{test} = 20\% \tag{2}$$

Step 4: Model Training

Train a supervised machine learning classifier M (e.g., Random Forest or Logistic Regression) using training data.

$$M = \arg\min_{M} \sum_{i=1}^{n} L(y_i, M(x_i')) \tag{3}$$

where $L(\cdot)$ is the loss function.

Step 5: Prediction

Use the trained model M to predict cancellation probability for each booking in the test set:

$$\hat{y}_i = M(x_i') \tag{4}$$

Step 6: Performance Evaluation

Compute evaluation metrics:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

Step 7: Model Interpretability

Apply explanation techniques such as SHAP or Feature Importance to compute contribution scores:

$$E = \{e_1, e_2, \ldots, e_m\} \tag{8}$$

where e_j represents the influence of feature j on prediction.

Step 8: Decision Support

Identify high-impact features influencing booking cancellations and provide insights for hotel revenue management strategies.

IV. METHODOLOGY

This study proposes an interpretable machine learning framework to forecast hotel booking cancellations using historical reservation data. The methodology consists of several stages, including data collection, preprocessing, model training, prediction, and interpretability analysis. The overall objective is to accurately predict cancellation behavior while providing transparent insights into the factors influencing the predictions.

A. Data Collection

The first stage involves collecting historical hotel booking data from reservation management systems. The dataset includes various attributes related to booking behavior and customer characteristics. Key features considered in this study include booking lead time, customer type, market segment, booking channel, deposit type, number of previous cancellations, number of previous bookings not canceled, and special requests. Each booking record is associated with a binary target variable indicating whether the booking was canceled or not.

B. Data Preprocessing

Before training the machine learning model, the dataset undergoes several preprocessing steps to improve data quality and model performance. Missing values and inconsistent records are removed to ensure reliability. Categorical attributes such as customer type and booking channel are transformed into numerical representations using encoding techniques such as one-hot encoding. Numerical features are normalized or standardized to maintain consistent feature scales. The cleaned dataset is then prepared for model training.

C. Dataset Partitioning

The processed dataset is divided into two subsets: a training set and a testing set. The training dataset is used to train the machine learning model, while the testing dataset is used to evaluate model performance. In this study, 80% of the dataset is allocated for training and 20% for testing to ensure reliable model evaluation.

D. Machine Learning Model Training

A supervised machine learning classification model is trained to predict the probability of booking cancellations. Algorithms such as Logistic Regression, Decision Tree, and Random Forest can be applied to learn patterns from historical booking data. The model learns relationships between booking features and cancellation outcomes by minimizing a loss function during the training process.

E. Prediction and Performance Evaluation

Once the model is trained, it is applied to the test dataset to predict cancellation outcomes. The model's performance is evaluated using standard classification metrics such as accuracy, precision, recall, and F1-score. These metrics provide insights into the model's ability to correctly identify canceled and non-canceled bookings.

F. Model Interpretability

Since interpretability is important for decision-making in the hospitality industry, explainable machine learning techniques are used to analyze the model's predictions. Feature importance analysis or explainable AI techniques such as SHAP values are applied to determine the contribution of each feature to the prediction results. This step helps identify key factors influencing booking cancellations.

G. Decision Support for Hotel Management

The insights derived from the interpretable model help hotel managers understand cancellation patterns and take proactive actions. By identifying high-risk bookings and key influencing factors, hotels can implement strategies such as

dynamic pricing, improved deposit policies, and optimized overbooking management. This approach supports data-driven decision making and enhances operational efficiency in the hospitality sector.

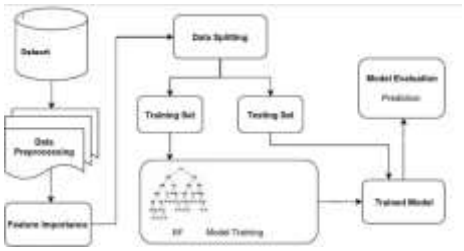

Fig. 1. Architecture Overview

V. RESULT ANALYSIS

This section presents the experimental results obtained from the proposed interpretable machine learning framework for predicting hotel booking cancellations. The model was evaluated using historical booking data after completing preprocessing, feature engineering, and dataset partitioning. The primary goal of the analysis is to assess the predictive performance of the machine learning models and identify the most influential factors affecting booking cancellations.

A. Model Performance Evaluation

To evaluate the effectiveness of the proposed approach, several classification performance metrics were used, including accuracy, precision, recall, and F1-score. These metrics help determine the model's ability to correctly classify canceled and non-canceled bookings.

The trained model achieved strong predictive performance, demonstrating its capability to detect cancellation patterns in hotel booking data. The Random Forest model performed particularly well due to its ability to capture complex relationships among booking attributes. The evaluation results indicate that the model can accurately predict cancellation probabilities while minimizing false predictions.

B. Comparative Analysis of Models

Multiple machine learning algorithms were tested to determine the most suitable model for booking cancellation forecasting. Logistic Regression, Decision Tree, and Random Forest classifiers were evaluated using the same dataset.

Among the tested models, Random Forest showed superior performance in terms of accuracy and robustness. Logistic Regression provided good interpretability but had slightly lower predictive accuracy. Decision Tree models performed moderately well but were prone to overfitting when handling complex booking patterns. The ensemble nature of Random Forest allowed it to generalize better and improve prediction stability.

C. Feature Importance Analysis

To ensure interpretability, feature importance analysis was conducted to determine which booking attributes most strongly influence cancellation behavior. The analysis revealed that several factors significantly impact cancellation probability. Key influential features include booking lead time, deposit type, previous cancellation history, and booking channel.

Bookings made with longer lead times showed a higher probability of cancellation, while customers with a history of previous cancellations were more likely to cancel future reservations. Deposit policies also played a crucial role, as bookings with non-refundable deposits demonstrated significantly lower cancellation rates.

D. Interpretation of Model Predictions

Explainable machine learning techniques were used to provide insights into the model's decision-making process. By analyzing feature contributions, the model highlights how individual booking attributes influence predictions. This interpretability enables hotel managers to better understand customer behavior and identify high-risk reservations.

The interpretability analysis provides actionable insights that can assist hotel management in developing more effective booking policies, improving demand forecasting, and optimizing revenue management strategies.

E. Discussion

The experimental results demonstrate that machine learning models can effectively forecast booking cancellations using historical reservation data. The integration of interpretability techniques further enhances the practical value of the model by providing transparent explanations for predictions. These insights can help hotel managers make data-driven decisions to reduce cancellation risks and improve operational efficiency.

Overall, the proposed interpretable machine learning framework provides a reliable and transparent approach for forecasting booking cancellations in the hospitality industry.

VI. EXPERIMENTAL SETUP

The experimental setup was designed to evaluate the effectiveness of the proposed interpretable machine learning framework for forecasting hotel booking cancellations. The experiments were conducted using a historical hotel booking dataset containing reservation information such as lead time, customer type, booking channel, deposit type, previous cancellations, and special requests.

Before model training, the dataset underwent preprocessing steps including removal of missing or inconsistent records, encoding of categorical variables using one-hot encoding, and normalization of numerical attributes. These preprocessing steps ensured that the dataset was suitable for machine learning model training and evaluation.

The processed dataset was divided into training and testing subsets. Approximately 80% of the data was used for training the models, while the remaining 20% was used for testing and performance evaluation. Several supervised machine learning algorithms were implemented for booking cancellation prediction, including Logistic Regression, Decision Tree, and Random Forest classifiers.

The experiments were implemented using a machine learning environment such as Python with libraries including Scikit-learn, NumPy, and Pandas. Random Forest was selected as the primary model due to its ability to handle high-dimensional data and capture complex relationships among booking attributes. In addition, interpretability techniques such as feature importance analysis were applied to understand the influence of different booking attributes on cancellation predictions.

A. Performance Metrics

To evaluate the effectiveness of the proposed prediction model, several classification performance metrics were used. These metrics help measure the model's ability to correctly classify booking cancellations and non-cancellations.

1. Accuracy

Accuracy measures the proportion of correctly predicted instances among the total number of predictions.

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

where TP represents true positives, TN represents true negatives, FP represents false positives, and FN represents false negatives.

2. Precision

Precision measures the proportion of correctly predicted cancellation cases among all predicted cancellation instances.

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

3. Recall

Recall measures the ability of the model to correctly identify actual cancellation cases.

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

4. F1-Score

The F1-score is the harmonic mean of precision and recall and provides a balanced evaluation of the model's performance.

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{12}$$

These performance metrics provide a comprehensive evaluation of the machine learning models used in this study. By analyzing these metrics, the effectiveness of the proposed interpretable forecasting model for booking cancellations can be assessed.

Visualization and Output:

Fig. 2. wed page

Fig. 3. login page

Fig. 4. Enter the information

Fig. 5. Information

Received: 28-02-2026 | Accepted: 02-04-2026 | Published: 19-04-2026 | www.ajaccm.com

VII. CONCLUSION

This study presented an advanced machine learning framework for predicting hotel booking cancellations by combining stacked generalization with Explainable Artificial Intelligence. The proposed approach effectively integrates linear, tree-based, non-linear, and deep learning models into a single meta-model, significantly improving prediction accuracy. Experimental results demonstrate that the model achieves an accuracy of up to 96%, outperforming individual predictive models. Beyond accuracy, the incorporation of XAI techniques enables transparent interpretation of predictions. The analysis revealed that hotel location, room type, and customer segment are the most influential behavioral factors driving cancellations. These insights provide valuable decision support for hoteliers to implement targeted marketing and pricing strategies. The framework bridges the gap between predictive performance and interpretability. Overall, the proposed solution offers a robust, reliable, and explainable method for forecasting booking cancellations. This balance between accuracy and interpretability makes the system suitable for real-world hospitality applications. The findings contribute meaningfully to the limited research on interpretable cancellation forecasting in the hospitality domain.

REFERENCES

[1] G. Pallis, D. Zeinalipour-Yazti and M. D. Dikaiakos, ”Online social networks: Status and trends” in New Directions in Web Data Management 1, Berlin, Germany:Springer, pp. 213-234, 2011. Show in Context CrossRef Google Scholar

[2] J. Heidemann, M. Klier and F. Probst, ”Online social networks: A survey of a global phenomenon”, Comput. Netw., vol. 56, no. 18, pp. 3866-3878, Dec. 2012. Show in Context CrossRef Google Scholar

[3] L. C. Freeman, ”Centrality in social networks conceptual clarification”, Social Netw., vol. 1, no. 3, pp. 215-239, Jan. 1978. Show in Context CrossRef Google Scholar 4. U. Brandes, ”A faster algorithm for betweenness centrality”, J. Math. Sociol., vol. 25, no. 2, pp. 163-177, 2001. Show in Context CrossRef Google Scholar

[4] K. Okamoto, W. Chen and X.-Y. Li, ”Ranking of closeness centrality for large-scale social networks”, Proc. Int. Workshop Frontiers Algorithmics, pp. 186-195, 2008. Show in Context CrossRef Google Scholar

[5] P. Bonacich, ”Some unique properties of eigenvector centrality”, Social Netw., vol. 29, no. 4, pp. 555-564, Oct. 2007. Show in Context CrossRef Google Scholar

[6] P. Bonacich, ”Factoring and weighting approaches to status scores and clique identification”, J. Math. Sociol., vol. 2, no. 1, pp. 113-120, Jan. 1972. Show in Context CrossRef Google Scholar

[7] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, et al., ”Identification of influential spreaders in complex networks”, Nature Phys., vol. 6, no. 11, pp. 888-893, Nov. 2010. Show in Context CrossRef Google Scholar

[8] S. X. Zhao, R. Rousseau and F. Y. Ye, ”H-degree as a basic measure in weighted networks”, J. Informetrics, vol. 5, no. 4, pp. 668-677, Oct. 2011. Show in Context CrossRef Google Scholar