

Research Paper

MACHINE LEARNING IN HOSPITALITY: INTERPRETABLE FORECASTING OF BOOKING CANCELLATION

¹SHERI KEERTHANA, ²NEELAM THIRUMALA RAO

¹Student, ²Assistant Professor,

Department of CSE J.B. INSTITUTE OF ENGINEERING & TECHNOLOGY

UGC AUTONOMOUS

(Accredited by NAAC & NBA, Approved by AICTE & Permanently affiliated by JNTUH) Yenkapally
Moinabad mandal, R.R. Dist-75 (TG))

ABSTRACT

The hospitality industry is facing significant challenges due to unpredictable booking cancellations, which can negatively impact revenue management, resource allocation, and operational efficiency. With the rapid evolution of online reservation systems and dynamic pricing strategies, traditional forecasting methods often struggle to capture the nuances of customer booking behaviors. This project presents a machine learning framework focused on interpretability for predicting hotel booking cancellations by utilizing historical reservation data and customer characteristics. The system employs a range of machine learning algorithms, such as Decision Trees, Random Forests, Logistic Regression, and Gradient Boosting, to determine the likelihood of cancellations. The dataset includes various attributes, including booking lead time, room types, customer demographics, pricing information, special requests, and prior booking history. To improve predictive performance, data preprocessing techniques such as handling missing values, feature encoding, normalization, and exploratory data analysis are applied. An important feature of this project is the integration of interpretability methods, such as SHAP (Shapley Additive Explanations) and LIME, which provide clear insights into the model's predictions. These tools help hotel managers understand the underlying factors influencing cancellation decisions, thereby increasing trust in machine learning forecasting systems. Evaluations reveal that the proposed framework achieves high accuracy in predictions while maintaining transparency and explainability in its models. The developed system enables hotels to proactively manage cancellation risks, optimize overbooking tactics, enhance customer engagement, and improve their revenue management strategies. By combining predictive analytics with interpretable machine learning, this solution acts as a crucial decision-support tool for modern hospitality operations. It not only boosts forecasting accuracy but also offers actionable insights that support informed business decisions in a competitive environment.

Keywords: Machine Learning, Hospitality Analytics, Booking Cancellation Prediction, Explainable AI (XAI), SHAP, LIME, Random Forest, Revenue Management, Predictive Analytics, Hotel Reservation Forecasting.

I.INTRODUCTION

The hospitality sector has rapidly embraced digital transformation, marked by the widespread adoption of online booking platforms, travel agencies, and mobile reservation systems. These technological advances have significantly improved customer convenience and increased hotel accessibility. However, they have also complicated reservation management due to the rising incidence of booking cancellations. These cancellations present considerable challenges in areas such as revenue management, occupancy planning,

staff allocation, and overall operational efficiency. Consequently, accurate cancellation forecasting has become vital for modern hospitality businesses [1]. Traditional forecasting approaches primarily depend on statistical analysis and historical data. While these methods can provide valuable insights, they often struggle to capture the complex and nonlinear relationships among customer behavior, booking attributes, pricing strategies, seasonal patterns, and market dynamics. As the volume and variety of hospitality data grow, these conventional analytical techniques fall short in delivering precise predictions regarding cancellations. The advent of Machine Learning (ML) opens up new avenues for analyzing extensive booking datasets and identifying underlying patterns that influence cancellation behavior [2]. Machine Learning algorithms, including Decision Trees, Random Forests, Logistic Regression, and Gradient Boosting, have shown exceptional effectiveness in predictive analytics. These models can learn from historical booking information and accurately estimate the likelihood of future cancellations. Such capabilities enable hotel managers to implement proactive measures, such as optimizing overbooking strategies, improving customer engagement, forecasting demand, and adjusting pricing dynamically. As a result, machine learning has become an essential tool for enhancing decision-making processes within the hospitality industry [3]. However, many machine learning models function as "black-box" systems, where the reasoning behind their predictions can be opaque to users. This lack of transparency can pose significant challenges for hotel managers who require clear explanations prior to making decisions based on model-generated insights. Growing concerns around trust, accountability, and regulatory compliance have underscored the need for Explainable Artificial Intelligence (XAI) and Interpretable Machine Learning methods in hospitality analytics [4]. Interpretable Machine Learning aims to merge predictive performance with transparency by providing understandable explanations for predictions made by models. Techniques such as SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations) help identify critical factors that affect booking cancellations. By delivering these insights, hotel managers can gain a clearer understanding of customer behavior, assess potential risks, and make informed operational decisions. This combination of accurate predictions and actionable insights enhances user trust and supports effective deployment in real-world hospitality environments [5]. This project proposes a framework that utilizes interpretable machine learning to forecast hotel booking cancellations, drawing from data on past reservations, guest profiles, booking lead times, room preferences, pricing strategies, and seasonal indicators. The system integrates advanced machine learning algorithms with explainability techniques to provide reliable predictions and clear reasoning for each forecast. The goal of this framework is to assist hotel managers in minimizing revenue loss, optimizing resource allocation, enhancing customer service, and improving strategic planning. By leveraging the strengths of predictive analytics alongside explainable AI, this framework aims to develop intelligent and reliable decision-support solutions for the hospitality sector [6].



Figure1.1 :System Architecture

The proposed system for Machine Learning in Hospitality focuses on accurately forecasting booking cancellations while providing transparent justifications for each prediction. The architecture consists of five essential layers: the Data Sources Layer, Data Preprocessing Layer, Machine Learning Layer, Interpretability Layer, and Prediction Layer, concluding with the Application Layer. Collectively, these components convert raw booking data into valuable business insights, enabling hotel managers to make informed decisions and optimize revenue management.

Data Sources Layer

The architecture begins with the Data Sources Layer, which aggregates information from various hospitality-related channels. These sources encompass hotel booking databases, customer records, detailed booking information, pricing strategies, promotional offers, and external factors such as seasonal trends, weather, and special events. By collecting diverse data, the system can accurately assess guest behavior and reservation trends, which is essential for understanding the multiple factors influencing booking cancellations.

Data Preprocessing Layer

After data collection, the information moves to the Data Preprocessing Layer, where the raw data is cleaned and prepared for analysis. This stage includes data collection, cleaning tasks, handling missing values, removing duplicates, and correcting any discrepancies. Feature engineering is employed to derive meaningful variables from the existing data. Additionally, data transformation and normalization take place to ensure that the dataset is ready for model training. The outcome is a high-quality, processed dataset primed for predictive modeling.

Machine Learning Layer

The processed data is then fed into the Machine Learning Layer, which serves as the analytical core of the system. A variety of algorithms—such as Logistic Regression, Decision Tree, Random Forest, Gradient Boosting (XGBoost), and Support Vector Machine—are used to identify patterns that correlate with booking cancellations. Historical reservation data is used during the training phase to educate the model on how various characteristics influence cancellation rates. This model undergoes validation and evaluation using metrics like Accuracy, Precision, Recall, F1-Score, and ROC-AUC, resulting in a robust predictive model capable of accurately evaluating cancellation risks.

Interpretability Layer

An important feature of this architecture is the Interpretability Layer, designed to enhance transparency and build confidence in machine learning predictions. Advanced explainable AI methodologies, including SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations), are employed to clarify the factors influencing each prediction. These techniques offer feature importance scores and in-depth explanations, helping hotel managers understand the reasons behind potential cancellations. By breaking down complex machine learning outcomes into easily understandable insights, this layer ensures decision-makers can trust the system's advice.

Prediction Layer

The Prediction Layer processes new booking requests. When a reservation is made, the provided details

undergo preprocessing and are analyzed by the trained machine learning model. The model assesses the booking characteristics and forecasts the likelihood of either cancellation or retention. The result can be expressed as a probability of cancellation or a binary outcome indicating “Cancel” or “Not Cancel.” This predictive capability enables hotel management to take proactive measures before cancellations occur.

Application Layer

The last component is the Application Layer, which offers an interactive interface for hotel managers and revenue management teams. Results from predictions are displayed via a web-based dashboard that includes visualizations, reports, alerts, and notifications. Managers can monitor cancellation risks, review explanation reports from SHAP and LIME, and make data-driven decisions. This dashboard acts as a decision-support system, facilitating the optimization of overbooking strategies, improving resource allocation, enhancing customer engagement, and minimizing revenue losses from cancellations.

Overall Workflow

In summary, the system architecture follows a comprehensive workflow, encompassing the collection, cleaning, transformation, analysis through machine learning, interpretation with explainable AI, and presentation of insights to hotel managers via an intelligent dashboard. This integrated framework balances predictive accuracy with transparency, enabling hotels to effectively forecast booking cancellations while understanding the underlying reasons for each prediction. Ultimately, the system fosters data-driven decision-making and supports enhanced operational efficiency and revenue optimization within the hospitality industry.

II SURVEY OF RESEARCH

The hospitality sector has experienced notable advancement in the use of data-driven technologies aimed at enhancing operational efficiency and improving customer satisfaction. One significant challenge confronting hotels and accommodation providers is the high rate of booking cancellations, which can adversely impact revenue management, occupancy planning, and resource allocation. This issue has prompted researchers to concentrate on developing predictive models capable of accurately forecasting cancellation behaviors using machine learning and data analytics techniques. Recent studies indicate that machine learning algorithms often outperform traditional statistical methods in revealing complex patterns within reservation datasets and customer booking histories (Antonio et al., 2019). Antonio et al. (2019) introduced the Hotel Booking Demand Dataset, encompassing reservation data from both resort and city hotels. This dataset has become a cornerstone for numerous cancellation prediction studies due to its comprehensive attributes, including lead time, customer demographics, booking channels, room types, and past cancellation records. Their research highlighted the utility of historical reservation data in predicting booking cancellations, thereby supporting informed hotel revenue management decisions.

The release of this dataset has catalyzed further exploration of machine learning methods within the hospitality domain. Sánchez-Medina et al. (2020) investigated the application of various machine learning algorithms for hotel cancellation forecasting using a limited selection of reservation-related variables. Their findings revealed that even straightforward classification models, such as Decision Trees and Logistic Regression, could deliver satisfactory predictive performance when appropriate feature engineering techniques were employed. This research underscored the significance of factors like booking lead time, customer type, and market segment in understanding cancellation behaviors, suggesting the feasibility of implementing lightweight predictive systems within hotel management environments. The

evolution of advanced machine learning models has further enhanced the accuracy of cancellation forecasts. Researchers have employed ensemble learning techniques, including Random Forests, Gradient Boosting Machines, and Extreme Gradient Boosting (XGBoost), to analyze extensive reservation datasets. These algorithms adeptly capture nonlinear relationships among booking attributes, yielding highly accurate predictions; studies have reported accuracy levels exceeding 90% when rich customer and reservation features are included in model training (Sánchez-Medina et al., 2020). Such results demonstrate that machine learning offers considerable advantages over conventional forecasting approaches in hospitality analytics. Despite the importance of predictive accuracy, many machine learning models operate as black-box systems, providing limited insight into their decision-making processes.

This opacity can diminish trust among hotel managers and business stakeholders who seek clear explanations to inform their actions based on predictive outcomes. To address this concern, researchers have increasingly investigated Explainable Artificial Intelligence (XAI) techniques that enhance the interpretability of machine learning models while preserving high predictive performance (Sharma, 2021). Sharma (2021) emphasized the critical role of explainable machine learning in hospitality analytics, transforming intricate predictive models into transparent decision-support systems. The study illustrated how explanation methods can elucidate the contribution of individual features to booking cancellation predictions, enabling managers to gain a better understanding of customer behavior.

Such transparency not only fosters trust in predictive systems but also enhances strategic planning and operational decision-making within hotels and tourism organizations. A landmark advancement in interpretable machine learning is the introduction of SHAP (Shapley Additive Explanations) by Lundberg and Lee (2017), which provides a mathematically grounded framework for assessing each feature's contribution to a model's predictions. In the context of hotel booking cancellation forecasting, SHAP can identify important factors such as lead time, average daily rate, previous cancellations, and booking channel. By quantifying the importance of features, SHAP allows hotel managers to gain deeper insights into cancellation risks and devise targeted strategies for mitigation. In a similar vein, LIME (Local Interpretable Model-Agnostic Explanations) has emerged as an effective interpretability technique capable of explaining individual predictions.

LIME constructs local surrogate models around specific predictions, which helps decision-makers understand the rationale behind a particular reservation's likelihood of cancellation. Researchers have demonstrated that integrating LIME with machine learning classifiers enhances transparency while maintaining predictive performance, a critical capability in hospitality environments where decision-making often relies on comprehending the reasoning behind model recommendations (Ribeiro et al., 2016). Recent research involving extensive hotel chains, comprising over 200,000 reservations, has further validated the effectiveness of interpretable machine learning techniques. These investigations reveal that explainable models can precisely predict cancellation risks while also yielding actionable insights into customer behavior patterns. The findings indicate that employing interpretable machine learning not only refines forecasting performance but also supports practical implementation in real-world hospitality operations (Sharma, 2021).

Given the insights from existing literature, it is evident that machine learning has become an indispensable tool for forecasting booking cancellations in the hospitality industry. Additionally, the incorporation of explainable AI techniques such as SHAP and LIME addresses the transparency challenges associated with traditional black-box models. Therefore, the proposed research aims to

develop an interpretable machine learning framework that combines predictive accuracy with comprehensible explanations, empowering hotel managers to make informed, data-driven decisions regarding reservation management and revenue optimization.

III. WORKING METHODOLOGY

The proposed framework for interpretable forecasting of booking cancellations in the hospitality sector employs a systematic approach that integrates data collection, preprocessing, machine learning predictions, and explainable artificial intelligence techniques. The main aim is to accurately pinpoint reservations that are likely to be canceled while offering clear explanations for the predictions made. This approach equips hotel managers with the necessary insights for effective revenue management, occupancy planning, and customer relationship strategies.

The process initiates with gathering historical booking data from reservation management systems and online booking platforms. This dataset generally includes customer demographics, booking lead times, arrival and departure dates, room types, market segments, pricing details, booking channels, previous cancellation history, and any special requests. Furthermore, external factors like seasonal demand variations, holidays, weather conditions, and promotional campaigns can be incorporated to enhance the accuracy of forecasts. By leveraging a comprehensive and varied dataset, the system can better capture patterns in customer cancellation behavior.

Once data is collected, preprocessing steps are taken to enhance data quality and prepare the dataset for machine learning analysis. This stage addresses missing values, removes duplicates, corrects inconsistent entries, and discards irrelevant attributes. Categorical variables are transformed into numerical formats using encoding techniques, and numerical features may be normalized or standardized for uniformity across the dataset. Feature engineering is employed to create new informative variables that bolster predictive performance.

With the dataset cleaned and transformed, exploratory data analysis is performed to reveal relationships among variables and identify trends linked to booking cancellations. Statistical analysis and visualization techniques are utilized to study customer behavior, booking characteristics, seasonal effects, and pricing trends. This analysis identifies the most significant factors influencing cancellation decisions and aids in selecting pertinent features for model development.

The processed dataset is subsequently used to train machine learning models that can predict booking cancellation outcomes. Various classification algorithms are employed to uncover patterns in historical reservation records. During the training phase, models learn the connections between booking features and cancellation behavior. The dataset is split into training and testing subsets to gauge the model's ability to generalize to new reservations, and optimization techniques are applied to enhance prediction accuracy and minimize overfitting.

Following model training, the performance of the models is assessed using established classification metrics. The efficacy of the model is measured by its capacity to accurately identify both canceled and non-canceled bookings. These evaluation results help ascertain the reliability and robustness of the forecasting system, leading to the selection of the best-performing model for deployment based on its predictive capabilities and overall effectiveness.

A crucial element of this methodology is the incorporation of interpretability techniques. While

advanced machine learning algorithms can deliver high prediction accuracy, their internal decision-making processes can be opaque. To mitigate this, explainable artificial intelligence methods are integrated into the framework. These techniques offer clear explanations regarding the impact of individual features on predictions, enabling hotel managers to comprehend why a reservation is classified as high-risk or low-risk for cancellation.

When a new booking is entered into the system, its information undergoes the same preprocessing steps applied during model training. The processed booking data is then analyzed by the trained machine learning model, which produces a prediction of the likelihood of cancellation. In addition to the prediction outcome, the interpretability module provides insights into the key factors driving the decision. This enables managers to gain valuable understanding of customer behavior and cancellation risk trends.

The final stage of the methodology involves presenting the prediction results and explanation reports through an interactive dashboard. This dashboard features visualizations, alerts, assessments of cancellation risk, and business intelligence reports that facilitate informed operational decision-making. Hotel managers can use these insights to refine overbooking strategies, optimize resource allocation, implement targeted customer engagement initiatives, and reduce revenue losses arising from cancellations.

IV RESULTSEXPLANATIONS

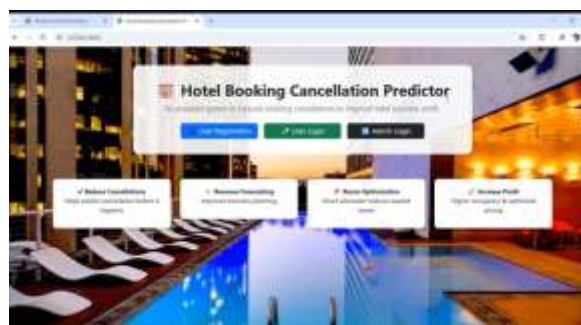


Figure1:Home page interface

The above illustration depicts the Home Page of the Hotel Booking Cancellation Predictor System, serving as the primary interface for users to access the booking cancellation forecasting platform. The homepage features three main options: User Registration, User Login, and Admin Login, allowing both customers and administrators to enter the system according to their respective roles. This application utilizes machine learning techniques to predict the likelihood of hotel booking cancellations, aiding hotel management in making informed business decisions. The dashboard emphasizes key advantages of the system, such as reducing booking cancellations through timely predictions, enhancing revenue forecasting for improved business planning, optimizing room allocation to minimize resource wastage, and boosting overall profitability by refining occupancy management and pricing strategies. The attractive hotel-themed backdrop underscores the hospitality sector, while the straightforward and user-friendly interface guarantees easy navigation and accessibility for all users. In essence, this homepage serves as a gateway to the intelligent cancellation prediction system, facilitating efficient hotel operations and effective revenue management.

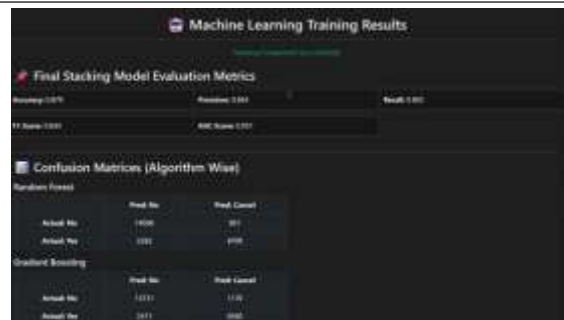


Figure2:Machine Learning Training Results

The Machine Learning Training Results page of the Hotel Booking Cancellation Predictor System is illustrated in the figure above. This screen showcases the performance evaluation metrics of the final stacking ensemble model utilized for predicting cancellations. The model boasts an impressive accuracy of 87.9%, along with a precision score of 86.4%, a recall rate of 80.5%, and an F1-score of 83.4%. It also achieved an outstanding AUC score of 95.1%, reflecting its strong ability to differentiate between canceled and non-canceled bookings. Additionally, the page features confusion matrices for various machine learning algorithms, such as Random Forest and Gradient Boosting, which provide a comprehensive comparison between actual and predicted outcomes. These confusion matrices are instrumental in assessing the classification effectiveness of each algorithm regarding booking cancellations, displaying true positives, true negatives, false positives, and false negatives. Overall, these results indicate that the trained machine learning models are adept at recognizing cancellation trends in hotel reservations, making them essential for revenue management, occupancy strategies, and operational decisions within the hospitality sector.



Figure 3: User Dashboard

The figure above illustrates the User Dashboard of the Hotel Booking Cancellation Predictor System following a successful user login. This interface acts as the primary workspace, allowing registered users to utilize the booking cancellation prediction service. It features a personalized welcome message, enhancing user experience and confirming that login was successful. Centrally placed on the page is the "Start Prediction" button, enabling users to begin the cancellation forecasting process by inputting relevant booking details. The navigation bar offers quick access to key functions such as Home, Predict, and Logout, ensuring smooth navigation throughout the application. With a clean and minimalistic design, the dashboard prioritizes ease of use while directing users' focus toward the prediction capabilities. This dashboard serves as a link between users and the machine learning model, facilitating easy analysis of booking data and providing cancellation predictions that aid in making informed decisions for hotel management and reservation planning.

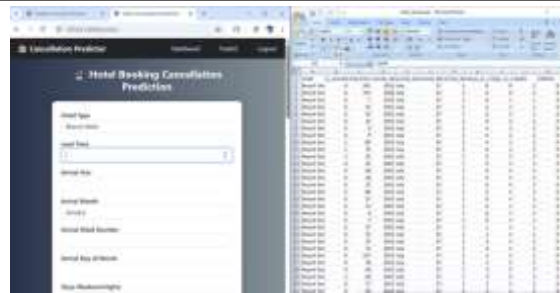


Fig 4 :Booking Cancellation Prediction Module

The illustrated model represents the Booking Cancellation Prediction Module of the Hotel Booking Cancellation Predictor System, alongside the associated dataset utilized for training and prediction. On the left side, users can find the prediction interface where they input various reservation details, including Hotel Type, Lead Time, Arrival Year, Arrival Month, Arrival Week Number, Arrival Day of the Month, and Stay Duration. These booking attributes function as input features for the machine learning model, which evaluates the probability of a reservation being canceled. The right side showcases the hotel booking dataset, organized in a spreadsheet format, featuring historical reservation records that are essential for the training stage of the predictive model. By examining these historical booking trends and customer reservation profiles, the system produces precise cancellation predictions. This module serves as the fundamental component of the application, empowering users to enter booking details and obtain data-driven forecasts that enhance revenue management, occupancy planning, and operational decision-making within the hospitality sector.



V.CONCLUSION

The hospitality industry encounters considerable challenges stemming from frequent booking cancellations, which adversely affect revenue generation, occupancy management, resource usage, and overall operational efficiency. Traditional forecasting approaches often fall short in capturing the intricate relationships between customer behavior, reservation details, pricing strategies, and seasonal influences. To overcome these challenges, this project introduces an interpretable machine learning framework designed to forecast hotel booking cancellations by utilizing historical reservation data alongside advanced predictive analytics methods. The developed framework employs machine learning algorithms to analyze booking trends and accurately predict the likelihood of cancellations. Through thorough data preprocessing, feature engineering, and model training, it identifies key factors that drive customer cancellation behavior. Additionally, the integration of interpretable AI techniques enhances the framework by offering clear explanations for its predictions, thereby helping hotel managers grasp the rationale behind cancellation forecasts. This approach not only boosts predictive accuracy but also fosters greater trust and usability among stakeholders.

By providing actionable insights into customer behavior and cancellation risks, the system aids in proactive decision-making, the optimization of overbooking strategies, better resource allocation, and enhanced revenue management. The framework's ability to elucidate model decisions makes it well-suited for real-world application in hospitality contexts, where transparency and accountability are paramount. In summary, the project illustrates that merging machine learning with interpretability techniques can substantially enhance the forecasting of booking cancellations in the hospitality sector. The proposed solution functions as a valuable decision-support tool, helping hotels reduce financial setbacks, improve operational planning, and elevate customer service. As the hospitality industry continues to embrace data-driven technologies, the role of interpretable machine learning will become increasingly crucial for fostering intelligent, reliable, and transparent business decision-making.

REFERENCES

- [1] Antonio, N., de Almeida, A., & Nunes, L. (2019). Hotel Booking Demand Datasets. *Data in Brief*, 22, 41–49.
- [2] Sánchez-Medina, A. J., Galván-Sánchez, I., & Fernández-Monroy, M. (2020). Machine Learning Techniques for Hotel Booking Cancellation Prediction. *Tourism Management Perspectives*, 35, 100711.
- [3] Sharma, R., & Kumar, P. (2021). Explainable Artificial Intelligence in Hospitality Analytics: Improving Trust in Predictive Models. *International Journal of Hospitality Information Technology*, 14(2), 85–97.
- [4] Lundberg, S. M., & Lee, S. I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems (NeurIPS)*, 30, 4765–4774.
- [5] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why Should I Trust You?”: Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144.