International Journal of Advanced Trends in Multidisciplinary Research
ISSN: 3139-0986, Volume 2 Issue 7
July 2026, PP 1-5

Machine Learning-Based Hotel Booking Cancellation Prediction and Visual Analytics Framework

S. Akhila[1], A. Ruchitha[2]
[1]Assistant Professor, Department of MCA, Megha Institute of Engineering and Technology for Women, Edulabad (Village), Ghatkesar (Mandal), Medchal District, Telangana – 501301, India.
[2]MCA Student, Department of MCA, Megha Institute of Engineering and Technology for Women, Edulabad (Village), Ghatkesar (Mandal), Medchal District, Telangana – 501301, India.

***Abstract* – Because they affect occupancy rates, revenue output, and operational planning, reservation cancellations are becoming an increasingly big issue for the hospitality industry. If hotel management can anticipate potential cancellations in advance of the scheduled check-in date, they may put effective reservation strategies in place and make informed decisions. This research presents a machine learning-based system for predicting hotel booking cancellations by assessing reservation-related variables such as booking lead time, guest data, room pricing, market segment, booking history, and special requests. The first step is to apply Exploratory Data Analysis (EDA) on the booking dataset to identify the factors that have a major influence on non-payment. Next, four supervised ML methods are trained and evaluated using various performance metrics, such as F1-score, recall, accuracy, and precision. These algorithms are Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree (DT), and Random Forest (RF). When compared to the other models, the Random Forest classifier performs the best in terms of prediction accuracy and total classification performance. Customers will find it much easier to understand the forecast results with the help of an interactive visual analytics tool that shows booking trends, cancellation details, and model comparisons. The proposed solution aids hotel management in a number of ways, including intelligent cancellation prediction, better reservation administration, and data-driven commercial decision-making. The decision-support capabilities of machine learning techniques have also been shown in related disciplines [1, 2].**

***Keywords* – Hotel Booking Cancellation, Machine Learning, Logistic Regression, Support Vector Machine, Decision Tree, Random Forest, Predictive Analytics, Visual Analytics, Reservation Management.**

I. INTRODUCTION

There has been tremendous expansion in the tourist and hospitality sector over the last few decades, which has boosted economies throughout the world and created a lot of new jobs. The hotel booking procedure has been made easier with the fast use of online reservation platforms and digital booking services. Customers can now make bookings effortlessly from anywhere. The downside is that hotel management is facing operational and financial issues due to the significant increase in cancellations caused by this convenience. The problem with frequent cancellations is that they make it hard to forecast room availability, diminish occupancy rates, waste resources, and lose money. A number of variables impact the likelihood of hotel reservation cancellations. These include the length of time before the booking, the room pricing, client preferences, travel dates, booking channels, and seasonal demand. The inability to resell empty rooms due to last-minute cancellations causes hotels to incur more promotional costs and see a decline in profit margins. As a result, finding out which bookings are most likely to be cancelled is now a must for better customer relationship management and company success.

The predictive analytics needs of many different types of businesses have been satisfactorily met by the latest developments in Machine Learning (ML). Machine learning algorithms are able to accurately predict the likelihood of future reservation cancellations by examining past booking data and discovering hidden trends. In order to optimise pricing strategies, enhance occupancy planning, and make educated business choices, hotel managers may use these predictive technologies to design proper cancellation policies.

Using data pertaining to reservations obtained from a publicly accessible dataset on hotel bookings, this study builds a framework based on machine learning to forecast when a reservation would be cancelled. To begin, the dataset is examined using Exploratory Data Analysis (EDA) to locate the most important features and comprehend client booking behaviour. Then, using common performance measures including accuracy, precision, recall, and F1-score, four supervised machine learning algorithms—Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree (DT), and Random Forest (RF)—are trained and assessed. In addition, a visual analytics tool is included to help customers better understand prediction findings by displaying booking patterns, cancellation information, and comparison model performance.

Below is a summary of the main points of this research: To find out what causes hotel reservations to be cancelled, we conduct a thorough Exploratory Data Analysis (EDA). Predictions of reservation cancellations are made using four different supervised machine learning models, which are then compared and tested. When evaluating prediction models, accuracy, precision, recall, and F1-score are used to choose the best one. With the use of an interactive interface, a web application based on visual analytics is created to examine booking information and aid hotel management in anticipating possible cancellations of reservations.

International Journal of Advanced Trends in Multidisciplinary Research
ISSN: 3139-0986, Volume 2 Issue 7
July 2026, PP 1-5

II. LITERATURE SURVEY

There has been a lot of focus on hotel booking cancellation prediction because of the big impact it has on revenue management, occupancy planning, and customer relationship management in the hospitality sector. Researchers have used a range of machine learning and statistical approaches to analyse booking trends and predict the likelihood of reservation cancellations, in response to the growing use of online reservation systems.

Prior research on the topic of hotel booking cancellation prediction has focused on classification algorithms including Decision Forests, Decision Trees, Support Vector Machines (SVM), Neural Networks, and ensemble learning techniques. These methods shown that hotels may enhance their room allocation tactics, decrease revenue losses, and back up successful overbooking choices by minimising the number of wrong forecasts. Because of their capacity to learn intricate correlations among booking variables, ensemble-based models reliably provided prediction performance among these methods.

Other researchers have used open-source hotel booking datasets and techniques including Logistic Regression, Random Forest, and Extreme Gradient Boosting (XGBoost). According to their research, Random Forest is better at classifying data since it deals with nonlinear connections and diverse booking information more successfully. Managers of hotels may optimise pricing strategies, boost client happiness, and increase operational efficiency with accurate cancellation prediction.

Aside from predictive modelling, review studies have brought attention to the increasing significance of AI and ML in the hotel industry. Demand forecasting, revenue management, recommendation systems, study of consumer behaviour, and dynamic pricing have all made good use of these technologies. Implementing these intelligent solutions successfully requires high-quality data, sufficient processing resources, and correct integration with current hotel management systems, despite the enormous advantages they bring.

Researchers have used statistical models to delve deeper into consumer booking habits and identify the elements that impact reservation cancellations. These factors include booking lead time, seasonal demand, market segment, customer geography, and group size. An increase in the frequency of booking cancellations is frequently seen to be associated with longer lead periods, online reservation channels, and seasonal changes. With this information, hotels can enhance their reservation planning and create cancellation policies that really work.

Predicting hotel booking cancellations effectively is still a tough endeavour, despite the encouraging findings reported in prior research. This is because booking behaviour is always changing based on consumer preferences and market circumstances. In light of that, this study builds a prediction framework based on machine learning by testing several models on data from hotel reservations, including Logistic Regression, Support Vector Machine, Decision Tree, and Random Forest. To further aid hotel managers in making well-informed operational choices, an interactive visual analytics tool is included to provide thorough insights into booking trends, model performance, and cancellation patterns.

Table I. Summary of Existing Research on Hotel Booking Cancellation

Reference	Machine Learning Technique	Major Findings
Study 1	Decision Forest, SVM, Neural Network	Reduced false predictions and improved booking management.
Study 2	Logistic Regression, Random Forest, XGBoost	Random Forest achieved superior cancellation prediction performance.
Study 3	Statistical and AI-based Analysis	Identified booking behavior and revenue management factors influencing cancellations.

III. PROPOSED SYSTEM

Through the application of supervised machine learning algorithms to customer reservation data, the suggested framework is able to forecast the likelihood of hotel booking cancellations. The whole process includes gathering data, cleaning it up, running exploratory analyses, building models, testing their efficacy, and finally, visualising the outcomes of the predictions. In order to make proactive business choices, hotel managers may use this methodical methodology to identify bookings that are likely to be cancelled.

The first step is to find a publically accessible source for the hotel booking dataset and check it for quality. Included in the dataset are details pertaining to reservations, such as the total number of adults and children, the type of room, the meal plan, the arrival details, the market segment, the price of the room, the prior bookings, the parking requirements, any special requests, and the booking status. Eliminating attributes that do not add value to the prediction process prepares the dataset for additional analysis without the need to fill in missing or duplicate values.

After that, we find out what the dataset is like and which variables have a big impact on booking cancellations by doing exploratory data analysis (EDA). Cancellation patterns, seasonal trends, client dispersion, booking behaviour, and hotel price are all examined using statistical summaries and graphical analytics. Insights into client reservation behaviour and assistance in selecting relevant features for model development are both provided by this analysis.

International Journal of Advanced Trends in Multidisciplinary Research
ISSN: 3139-0986, Volume 2 Issue 7
July 2026, PP 1-5

In order to maintain the class distribution of booking status, the dataset is stratified sampled after preprocessing and then split into training and testing subsets. To facilitate effective data processing by machine learning algorithms, suitable encoding methods are used to convert categorical features into numerical representations.

Four supervised machine learning techniques are used to forecast when a booking will be cancelled:

- Logistic Regression (LR)
- Support Vector Machine (SVM)
- Decision Tree (DT)
- Random Forest (RF)

Accuracy, precision, recall, and F1-score are some of the common classification measures used to assess each model's performance after training on the provided dataset. The best model for predicting possible reservation cancellations may be found by comparing these algorithms. The findings show that the Random Forest classifier can capture complicated correlations among booking variables, which gives it greater prediction capacity.

A web-based application with visual analytics support is used to incorporate the trained prediction model, which improves user engagement. Using the app's user-friendly interface, users may examine booking details, compare model performance, see patterns in cancellations, and get forecast outcomes. As a result, this helps hotel management reduce financial losses and maximise operational efficiency via the use of smart reservation tactics.

Fig-1. Heatmap of the correlation matrix

Table II. Machine Learning Models Used in the Proposed Framework

Model	**Purpose**	**Performance Evaluation**
Logistic Regression	Baseline classification model	Accuracy, Precision, Recall, F1-Score
Support Vector Machine	Linear and non-linear classification	Accuracy, Precision, Recall, F1-Score
Decision Tree	Rule-based prediction	Accuracy, Precision, Recall, F1-Score
Random Forest	Ensemble learning for robust prediction	Accuracy, Precision, Recall, F1-Score

IV. EXPERIMENTAL ANALYSIS

Using a publicly accessible dataset of hotel reservations, we test how well the suggested framework predicts the cancellation of hotel bookings. Encoding categorical characteristics and preparing features for supervised learning are part of the preprocessing steps that are performed on the acquired data before model training. Stratified sampling ensures that the distribution of booking status is preserved when the processed dataset is separated into training and testing subsets in a 70:30 ratio. This method guarantees that the cancelled and non-canceled reservation samples in both subgroups are statistically valid.

Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree (DT), and Random Forest (RF) are the four supervised machine learning techniques that make up the prediction framework. In order to provide an accurate picture of how well each algorithm predicts, they are all trained using the same data. The capacity of the trained models to detect hotel booking cancellations is next tested on the testing dataset.

Accuracy, Precision, Recall, and F1-Score are some of the most used classification metrics used to evaluate model performance. Precision assesses the accuracy of expected cancellations, whereas Accuracy examines the model's overall prediction capabilities. By integrating Precision and Recall, the F1-Score offers a well-rounded metric that assesses the model's capacity to detect real cancelled reservations. Each classifier is thoroughly evaluated using these measures.

According to the findings of the experiments, all four machine learning models do a good job of classifying hotel booking information. While Support Vector Machine shows better classification by defining optimum decision boundaries, Logistic Regression gives a solid foundation for binary classification. By using hierarchical decision rules, Decision Tree is able to capture the link between booking characteristics and cancellation status. The Random Forest classifier outperforms all other models in terms of prediction performance. It does this by integrating several decision trees, which helps to decrease overfitting and improves generalisability. So, to forecast hotel booking cancellations, Random Forest is chosen as the best model.
The prediction framework uses visual analytics to provide booking patterns, cancellation distribution, and comparison model performance, which further improves interpretability. Hotel managers may learn about guests' booking habits, see trends in cancellations, and make data-driven decisions about reservations with the help of these analytical visualisations.

V. RESULTS AND DISCUSSION

The experimental results show that by analysing reservation trends in the past, machine learning algorithms can accurately forecast when a hotel reservation will be

cancelled. Once the chosen classification models have been trained and tested, we evaluate their predicted performance using common assessment criteria like F1-Score, Accuracy, Precision, and Recall. By contrasting their learning capacities and decision-making strategies, we can see that the algorithms make distinct contributions to the prediction process.

As a solid foundational model for binary classification, Logistic Regression delivers respectable results. By building optimum decision boundaries, the Support Vector Machine enhances prediction and makes it easier to distinguish between reservations that have been cancelled and those that have not. By determining which booking attributes are most essential, the Decision Tree classifier can then provide competitive prediction results and decision rules that are easy to understand.

The Random Forest classifier outperforms all other models that were tested. To make predictions more stable and less prone to overfitting, it uses an ensemble learning method that merges several decision trees. With better generalisation to unseen booking data and enhanced prediction accuracy, the model effectively captures complicated interactions among booking variables. Thus, Random Forest is determined to be the optimal model for predicting the cancellation of hotel reservations.

Feature analysis goes a step further and shows that the most important factor in reservation cancellations is the booking lead time. Customers who prepare ahead are more likely to back out if their trip plans change or if they find a better choice. Average room pricing, market sector, special request number, arrival month, and booking history are other important considerations. The machine learning models' prediction power is enhanced by these features when taken as a whole.

By displaying cancellation patterns, booking distributions, and model comparisons via interactive visualisations, the created visual analytics module increases the usability of the proposed system. With this information, hotel managers may better comprehend client habits, enhance reservation management tactics, and take preventative actions to lessen financial losses caused by cancellations. To optimise hotel income and operational efficiency, the suggested framework offers a realistic and trustworthy decision-support system.

VI. WEB APPLICATION IMPLEMENTATION

The suggested system for predicting hotel booking cancellations is built as a web app using the Django framework. It gives hotel managers an easy-to-use platform to track reservation activity and anticipate possible cancellations. By combining the trained ML model with a user-friendly web interface, the Django framework allows for centralised data administration and real-time prediction. Through the app's user-friendly interface, users are able to input or upload hotel reservation details. Following the submission of the booking details, the Django backend processes the data and sends it to the trained ML model for prediction. The system then promptly notifies the user of its cancellation prediction if it thinks the reservation is likely to be cancelled.

Visual analytics are also a part of the program, and they provide crucial booking data, cancellation patterns, monthly reservation distributions, and comparison model performance, in addition to prediction. With the help of these analytical visualisations, hotel managers may better comprehend client booking habits, spot bookings with a high chance of cancellation, and make educated choices about daily operations. With the integrated dashboard, customers can easily keep tabs on booking activity and assess forecast results.

Thanks to its organised backend design, the deployed Django software also allows for safe data storage, maintenance of reservation records, and model integration. Reduced revenue losses, better reservation planning, and enhanced hotel management operations are all possible thanks to the suggested system's use of machine learning in conjunction with a web-based deployment platform. In addition, the application's modular design makes it easy to include new business intelligence tools and complex predictive models in the future without having to completely overhaul the current system.

VII. CONCLUSION

Using past reservation data and booking-related variables, this study introduces a system based on machine learning that can forecast hotel booking cancellations. In order to correctly categorise reservation status, the suggested system uses Exploratory Data Analysis (EDA) to examine client booking behaviour and then applies several supervised machine learning techniques, such as Logistic Regression, Support Vector Machine, Decision Tree, and Random Forest. The Random Forest classifier outperforms the others in terms of prediction performance, making it the best option for spotting possible booking cancellations, according to the comparison.

The trained model is linked into a web application that is deployed using the Django framework. This application allows hotel managers to do real-time cancellation prediction and see booking patterns via an interactive dashboard, making the prediction framework more practical and usable. By assisting with the identification of high-risk bookings, optimisation of room allocation, improvement of pricing strategies, and reduction of revenue losses due to unexpected cancellations, the system enables data-driven decision-making in the hotel industry.

The suggested framework integrates machine learning methods with an existing Django web platform to provide an intelligent solution for managing hotel reservations; it is

also scalable, dependable, and efficient. Among the many benefits to the hotel sector, the created system boosts forecast accuracy, operational planning, and customer service. To further increase prediction performance and scalability, future additions might include integrating deep learning models, real-time reservation data, and cloud-based deployment.

REFERENCES

1. N. Antonio, A. M. de Almeida, and L. Nunes, "Predicting Hotel Booking Cancellations to Decrease Uncertainty and Increase Revenue," Tourism & Management Studies, vol. 13, no. 2, pp. 25–39, 2017.
2. N. Antonio, A. de Almeida, and L. Nunes, "Predicting Hotel Bookings Cancellation with a Machine Learning Classification Model," in Proc. 16th IEEE Int. Conf. on Machine Learning and Applications (ICMLA), 2017, pp. 1049–1054.
3. N. Antonio, A. de Almeida, and L. Nunes, "An Automated Machine Learning Based Decision Support System to Predict Hotel Booking Cancellations," Data Science Journal, vol. 18, no. 1, 2019.
4. N. Antonio, A. de Almeida, and L. Nunes, "Using Machine Learning and Big Data for Efficient Forecasting of Hotel Booking Cancellations," International Journal of Hospitality Management, vol. 89, 2020.
5. M. Falk and M. Vieru, "Modelling the Cancellation Behaviour of Hotel Guests," International Journal of Contemporary Hospitality Management, vol. 30, no. 10, pp. 2948–2967, 2018.
6. E. Alotaibi, "Application of Machine Learning in the Hotel Industry: A Critical Review," Journal of Association of Arab Universities for Tourism and Hospitality, vol. 19, no. 3, 2020.
7. P. Kumar and S. Sharma, "Hotel Booking Prediction Using Machine Learning," International Journal for Research in Applied Science and Engineering Technology, vol. 10, no. 4, 2022.
8. Z. A. Andriawan, S. R. Purnama, A. S. Darmawan, R. Ricko, A. Wibowo, A. Sugiharto, and F. Wijayanto, "Prediction of Hotel Booking Cancellation Using CRISP-DM," in Proc. Int. Conf. on Informatics, Multimedia, Cyber and Information System, 2020.
9. E. C.-S. Agustín, J. Sánchez-Medina, and L. Romero-Domínguez, "Forecasting Hotel Booking Cancellations Using Personal Name Records: An Artificial Intelligence Approach," 2022.
10. S. Chen, E. W. T. Ngai, Y. Ku, Z. Xu, X. Gou, and C. Zhang, "Prediction of Hotel Booking Cancellations: Integration of Machine Learning and Probability Model Based on Interpretable Feature Interaction," Decision Support Systems, vol. 170, Art. 113959, 2023.
11. A. Herrera, "Forecasting Hotel Cancellations Through Machine Learning," Expert Systems, vol. 41, 2024.
12. M. A. Jishan, V. Singh, A. K. Ghosh, M. S. Alam, K. R. Mahmud, and B. Paul, "Hotel Booking Cancellation Prediction Using Applied Bayesian Models," 2024.
13. J. Sun, "Hotel Booking Cancellation and Machine Learning," Applied and Educational Intelligence, 2025.
14. P. Silvestre, N. Antonio, and P. Carrasco, "Navigating Uncertainty: Enhancing Hotel Cancellation Predictions with Adaptive Machine Learning," Information Technology & Tourism, vol. 28, 2026.
15. W. S. Chow, C. Parry, and J. C. Chan, "Developing a Forecast System for Hotel Occupancy Rate Using Integrated Forecasting Models," Journal of Hospitality and Tourism Management.
16. S. Chatterjee, "Drivers of Helpfulness of Online Hotel Reviews: A Sentiment and Emotion Mining Approach," International Journal of Hospitality Management, 2019.
17. C. M. Bishop, Pattern Recognition and Machine Learning. New York, NY, USA: Springer, 2006.
18. T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, 2nd ed. New York, NY, USA: Springer, 2009.
19. L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
20. C. Cortes and V. Vapnik, "Support-Vector Networks," Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.
21. P. Ramesh Babu, "An Improved Deep Learning Methodology Design to Predict the Presence and Stage of Knee Osteoarthritis Using X-Ray Image," in Proc. IEEE Int. Conf., Feb. 2026.
22. P. Ramesh Babu, "Experimental Evaluation to Detect Child Malnutrition Based on Pediatric Visual Data Assessment Using Deep Learning Procedure," in Proc. IEEE Int. Conf., Feb. 2026.