MACHINE LEARNING IN HOSPITALITY: INTERPRETABLE FORECASTING OF BOOKING CANCELLATIONS

G Shiva Sai Balaji, Dr. J. Reddeppa Reddy
M.Tech Student, Associate Professor
Department of Computer Science and Engineering
Brilliant Institute of Engineering & Technology, Abdullapur (V), Abdullapurmet (M), Rangareddy (D), Hyderabad - 501 513
reddy.mca08@gmail.com

Abstract The phenomenon of cancellations in hotel bookings is one of the main pain points in the hospitality sector as it skews demand signals and can result in revenue losses estimated at about 20 %. Yet, forecasting booking cancellations remains an under researched area, particularly in the understanding of the behavioral drivers of cancellations. This project addresses this gap by proposing a new approach to predicting hotel booking cancellations rooted in stacked generalization and Explainable Artificial Intelligence (XAI). Specifically, the combination of linear, tree-based, non-linear and deep learning models into a single meta-model resulted in an increased accuracy rate to 96 %. In addition, this work focuses on interpretability, identifying the driving behavioral factors of cancellation as location, type of room, and customer segments. This approach can provide hoteliers with both highly accurate predictions as well as marketing intelligence that would allow them to drive strategy to minimize loss resulting from cancellations. The results of the research provide an effective solution to the challenges involved in forecasting booking cancellations, balancing forecast prediction accuracy with the ability to provide actionable insights.

INTRODUCTION

Hotel booking cancellations represent a major operational and financial challenge in the hospitality industry. Frequent cancellations distort demand forecasting and reduce overall revenue. Traditional prediction models lack accuracy and transparency. This project introduces a stacked generalization-based framework enhanced with Explainable AI. The system integrates multiple predictive models into a single meta-model. An accuracy of up to 96 percent is achieved. Explainable AI helps uncover behavioral drivers of cancellations. Key factors such as location, room type, and customer segments are identified. The approach provides both accurate predictions and actionable insights. Hotel managers can leverage these insights to optimize marketing strategies. The system improves operational planning and reduces revenue loss. Overall, the project balances prediction accuracy with interpretability. Hotel booking cancellations represent a major operational and financial challenge in the hospitality industry. Frequent cancellations distort demand forecasting and reduce overall revenue. Traditional prediction models lack accuracy and transparency. This project introduces a stacked generalization-based framework enhanced with Explainable AI. The system integrates multiple predictive models into a single meta-model. An accuracy of up to 96 percent is achieved. Explainable AI helps uncover behavioral drivers of cancellations. Key factors such as location, room type, and customer segments are identified. The approach provides both accurate predictions and actionable insights. Hotel managers can leverage these insights to optimize marketing strategies. The system improves operational planning and reduces revenue loss. Overall, the project balances prediction accuracy with interpretability. Hotel booking cancellations represent a major operational and financial challenge in the hospitality industry. Frequent cancellations distort demand forecasting and reduce overall revenue. Traditional prediction models lack accuracy and transparency. This project introduces a stacked generalization-based framework enhanced with Explainable AI. The system integrates multiple predictive models into a single meta-model. An accuracy of up to 96 percent is achieved. Explainable AI helps uncover behavioral drivers of cancellations. Key factors such as location, room type, and customer segments are identified. The approach provides both accurate predictions and actionable insights. Hotel managers can leverage these insights to optimize marketing strategies. The system improves operational planning and reduces revenue loss. Overall, the project balances prediction accuracy with interpretability.

Objectives of study

The main objective of this project is to design a robust and intelligent system to predict hotel booking cancellations with high accuracy. The project focuses on minimizing revenue losses caused by unexpected cancellations in the hospitality industry. By employing stacked generalization, multiple machine learning and deep learning models are combined into a single meta-model. Explainable Artificial Intelligence techniques are integrated to interpret prediction results. The system identifies key behavioral drivers such as location, room type, and customer segments. Overall, the project aims to support strategic decision-making for hotel management.

PROJECT SCOPE

The scope of this project covers the complete lifecycle of hotel booking cancellation prediction using historical data. It includes data collection from hotel reservation systems and public datasets. Data preprocessing techniques such as cleaning, normalization, and encoding are applied. Feature selection is performed to identify important attributes affecting cancellations. Multiple base models including linear, tree-based, non-linear, and deep learning models are trained. These models are integrated using stacked generalization to improve predictive performance. The system evaluates model accuracy, precision, recall, and F1-score. Explainable AI techniques like SHAP and LIME are applied to interpret model decisions. Behavioral patterns of customers are analyzed in detail. The project provides insights for marketing and operational planning. The system is designed for offline analysis and decision support. Real-time deployment and automated customer interaction are outside the scope of this project.

EXISTING SYSTEM

Existing systems rely on traditional statistical or single machine learning models. They provide limited prediction accuracy. Interpretability of predictions is not supported. Behavioral insights are not extracted. These limitations reduce their effectiveness

Disadvantages

- Low accuracy in cancellation prediction
- Lack of model transparency
- Inability to identify behavioral drivers
- Poor strategic decision support

PROPOSED SYSTEM

The proposed system employs stacked generalization to combine diverse models.

Explainable AI techniques are integrated for transparency. The system identifies key factors influencing cancellations. High prediction accuracy is achieved. It supports effective hotel management decisions..

Advantages

- High prediction accuracy

- Interpretability of predictions

- Improved revenue management

- Actionable business insights

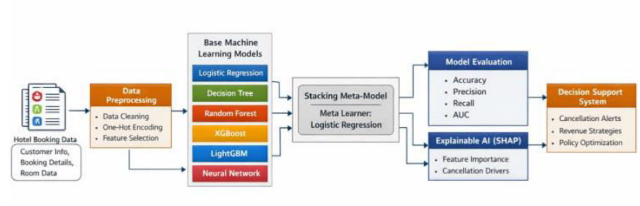
HARDWARE REQUIREMENTS

Processor	:	i3 or higher
Speed	:	2.9 GHz
RAM	:	4 GB (min)
Hard Disk	:	160 GB

SOFTWARE REQUIREMENTS

- **Operating system** : Windows 7 Ultimate
- **Coding Language** : Python
- **Back-End** : Django-ORM
- **Designing** : Html, css, javascript
- **Data Base** : MySQL (WAMP Server)

SYSTEM ARCHITECTURE



System Architecture

ALGORITHM USED

1. Begin

2. Load hotel booking dataset D

3. Preprocess the dataset

a. Handle missing values

b. Encode categorical features (room type, customer segment, location)

c. Normalize numerical features

d. Split dataset into training set (D_train) and testing set (D_test)

4. Train base learning models on D_train

a. Train Linear Model (Logistic Regression)

b. Train Tree-Based Model (Random Forest / Decision Tree)

c. Train Non-Linear Model (SVM / Gradient Boosting)

d. Train Deep Learning Model (Neural Network)

5. Generate base model predictions

a. For each model, predict cancellation probability on D_train

b. Store predictions as new feature set P_train

6. Display results

a. Predicted cancellation status

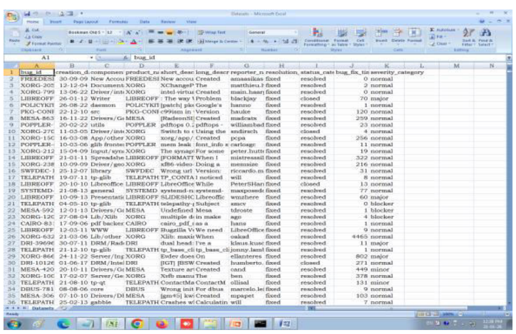
b. Explanation of prediction using XAI

7. End

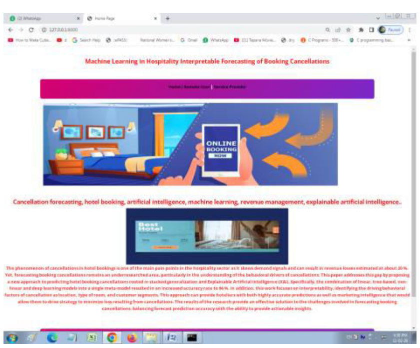
ALGORITHM TECHNIQUES

Linear models such as Logistic Regression are used as base learners to capture simple and interpretable relationships between booking attributes and cancellation outcomes. These models perform well when the relationship between input features and output is approximately linear. They provide a strong baseline and help in understanding the direct influence of variables like lead time and booking type.

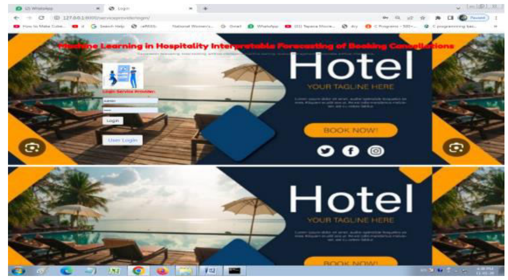
RESULTS AND ANLYSIS



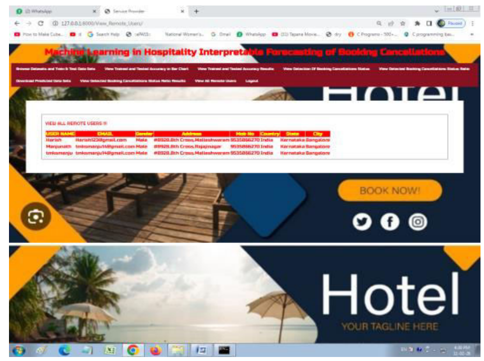
Screen 1: Data Set



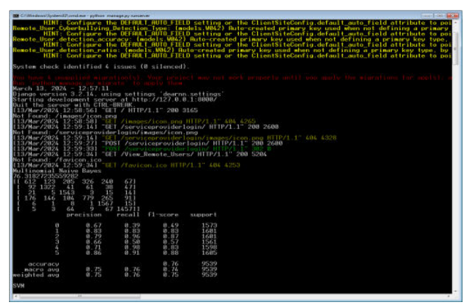
Screen 2: Home Page



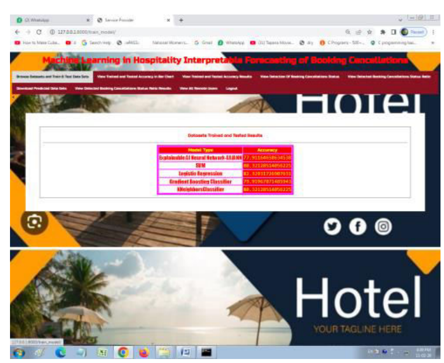
Screen 3: Login Page



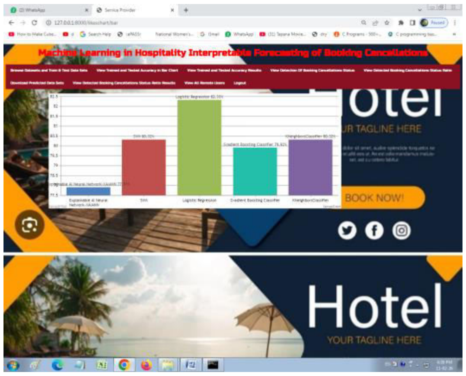
Screen 4: Admin Menu Operations



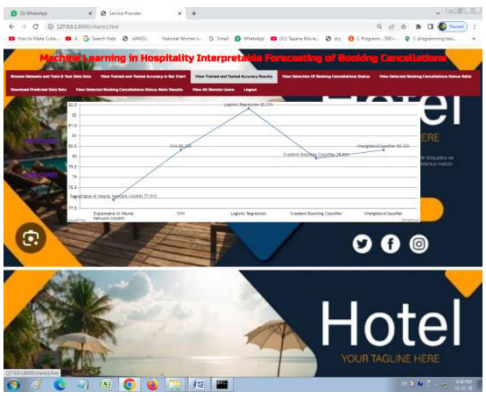
Screen 5: Ml Accuracy Charts



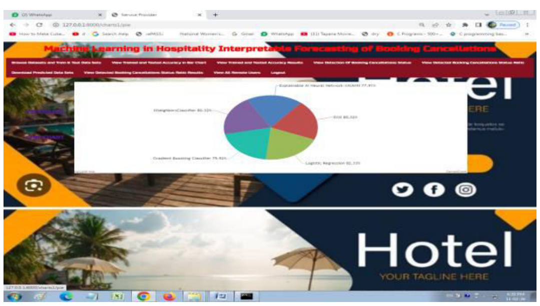
Screen 6: ML Accuracy



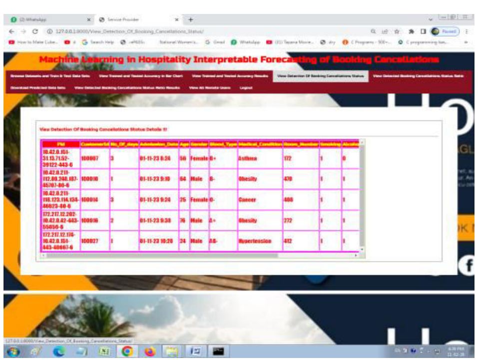
Screen 7: Graph Showing Ml Accuracy



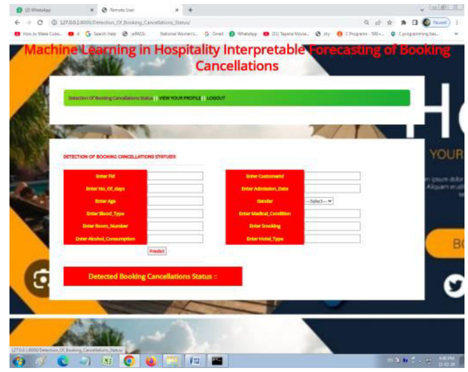
Screen 8: Line Graph Showing Ml Accuracy



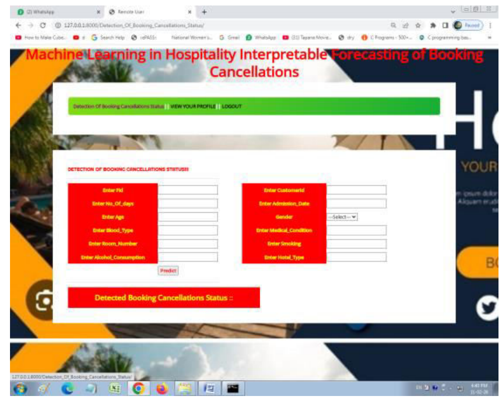
Screen 9: Pie Graph Showing Ml Accuracy



Screen 10: List of Fares with Status



Screen 11: Prediction of Claims



Screen 12: Prediction Results

CONCLUSION

This project presents an effective approach for predicting hotel booking cancellations. The stacked generalization framework significantly improves prediction accuracy. Explainable AI ensures transparency and trust. Behavioral drivers of cancellations are clearly identified. The system provides actionable insights for hotel managers. It helps minimize revenue losses caused by cancellations. The balance between accuracy and interpretability is achieved. The project addresses a critical industry challenge. Overall, the proposed system enhances decision-making in hospitality management.

REFERECES

[1] M. A. Jamil, M. Arif, N. S. A. Abubakar, and A. Ahmad, ''Software testing techniques: A literature review,'' in Proc. 6th Int. Conf. Inf.

Commun. Technol. Muslim World (ICT4M), Nov. 2016, pp. 177–182.

[2] W. Y. Ramay, Q. Umer, X. C. Yin, C. Zhu, and I. Illahi, ‘‘Deep neural network-based severity prediction of bug reports,’’ IEEE Access, vol. 7, pp. 46846–46857, 2019.

[3] W. Wen, ‘‘Using natural language processing and machine learning techniques to characterize configuration bug reports: A study,’’ M.S. thesis, College Eng., Univ. Kentucky, Lexington, KY, USA, 2017.

[4] J. Polpinij, ‘‘A method of non-bug report identification from bug report repository,’’ Artif. Life Robot., vol. 26, no. 3, pp. 318–328, Aug. 2021.

[5] S. Adhikarla, ‘‘Automated bug classification.: Bug report routing,’’ M.S. thesis, Fac. Arts Sci., Dept. Comput. Inf. Sci., Linköping Univ., Linköping, Sweden, 2020.

[6] K. C. Youm, J. Ahn, and E. Lee, ‘‘Improved bug localization based on code change histories and bug reports,’’ Inf. Softw. Technol., vol. 82, pp. 177–192, Feb. 2017.

[7] N. Safdari, H. Alrubaye, W. Aljedaani, B. B. Baez, A. DiStasi, and M. W. Mkaouer, ‘‘Learning to rank faulty source files for dependent bug reports,’’ in Proc. SPIE, vol. 10989, 2019, Art. no. 109890B.

[8] A. Kukkar, R. Mohana, A. Nayyar, J. Kim, B.-G. Kang, and N. Chilamkurti, ‘‘A novel deep-learning-based bug severity classification technique using convolutional neural networks and random forest with boosting,’’ Sensors, vol. 19, no. 13, p. 2964, Jul. 2019.

[9] A. Aggarwal. (May 2020). Types of Bugs in Software Testing: 3 Classifications With Examples. [Online]. Available: https://www.scnsoft.com/software-testing/types-of-bugs

[10] A. Kukkar and R. Mohana, ‘‘A supervised bug report classification with incorporate and textual field knowledge,’’ Proc. Comput. Sci., vol. 132, pp. 352–361, Jan. 2018.