Project documentation - Hotelier's Challenge!

by Overfitting Overlords

1. Introduction

1.1. Project Overview

This project focuses on developing predictive models to forecast the occupancy rates of hotel rooms. Utilizing historical data, the project comprises three multi-step prediction models that forecasts future room occupancy over varying intervals (day, week, month) and three one-step models that predict daily, weekly, and monthly occupancy rates. The integration of these models aims to enhance decision-making processes for hotel management, allowing for more efficient resource allocation, pricing strategies, and overall operational optimization.

1.2. Objectives

The primary objectives of this project are:

- **To develop a robust multi-step prediction model** that can forecast daily, weekly, and monthly number of occupied rooms in a hotel for future dates.
- **To create accurate one-step prediction models** for daily, weekly, and monthly hotel room occupancy.
- **To evaluate the performance of these models** in terms of accuracy and reliability, ensuring they are practical tools for operational planning.
- **To provide actionable insights** to hotel management on optimizing occupancy rates and improving revenue management.

1.3. Scope and Significance

The scope of this project extends to the application of advanced data analytics techniques in the hospitality industry, particularly in the context of occupancy prediction. The significance of this project lies in its potential to transform hotel operations by providing predictive insights that can lead to enhanced customer service, optimized revenue management, and improved operational efficiency. The predictive models developed in this project are intended to serve as a decision-support tool, guiding hotel managers in making informed decisions based on anticipated occupancy trends. Additionally, the methodologies and findings of this project could serve as a benchmark for similar applications in other sectors of the hospitality industry.

2. Methodology

2.1. Data Preprocessing

The preprocessing of the collected data involved several crucial steps to prepare it for effective modeling:

1. **Cleaning**: Removal of erroneous entries that could skew the model results, such as bookings that were canceled or unoccupied at atypical dates (e.g. reservation_date after stay_date).
2. **Transformation**: Conversion of categorical data into a numerical format suitable for analysis, using techniques such as one-hot encoding for attributes like room type.
3. **Feature Engineering**: Creation of new features that could potentially enhance model performance, such as deriving the day of the week from date records, or calculating historical occupancy rates during similar periods (**lag features**).

2.2. Model Selection and Justification

The primary objectives of this project are:

- **To develop a robust multi-step prediction model** that can forecast the number of occupied rooms in a hotel for future dates (daily, weekly and monthly).
- **To create accurate one-step prediction models** for daily, weekly, and monthly hotel room occupancy.
- **To evaluate the performance of these models** in terms of precision and reliability, ensuring they are practical tools for operational planning.
- **To provide actionable insights** to hotel management on optimizing occupancy rates and improving revenue management.

2.3. Scope and Significance

The scope of this project extends to the application of advanced data analytics techniques in the hospitality industry, particularly in the context of occupancy prediction. The significance of this project lies in its potential to transform hotel operations by providing predictive insights that can lead to enhanced customer service, optimized revenue management, and improved operational efficiency. The predictive models developed in this project are intended to serve as a decision-support tool, guiding hotel managers in making informed decisions based on anticipated occupancy trends. Additionally, the methodologies and findings of this project could serve as a benchmark for similar applications in other sectors of the hospitality industry.

3. Model description

3.1. Multi-Step Prediction Model for Occupied Rooms

The multistep forecasting models consist of three separate models: **multi_step_day**, **multi_step_week**, and **multi_step_month**. Each model is designed to address specific forecasting needs for hotel room bookings, predicting daily, weekly, and monthly booking volumes respectively. This approach allows for tailored strategies that optimize predictions according to the varying patterns observed at different time scales.

3.1.1. Models

- **multi_step_day**: Predicts the number of room bookings per day.
- **multi_step_week**: Forecasts the total number of room bookings for each week.
- **multi_step_month**: Estimates the total number of room bookings for each month.

3.1.2. Purpose

These models enable precise forecasting of room demand, aiding operational planning and revenue management for hotels by predicting future booking volumes.

3.1.3. Data Management

3.1.3.1. Data Segregation

- **training data**: each model is trained on one full year of historical room booking data, which includes dates and room counts
- **testing data**: another full year of data is used to test the models, ensuring that they are evaluated on their ability to generalize to unseen data

3.1.3.2. Data Handling Strategy

The approach avoids the use of recursive features where outputs of the model for previous time steps are fed back as inputs for future predictions. This strategy was adopted after identifying that models with lag features performed poorly, potentially due to the introduction of noise and overfitting risks.

3.1.4. Forecasting Approach

Non-Recursive Multi Step Forecasting: The term 'multistep' reflects the models' capability to forecast several steps into the future without relying on the immediate past outputs. This method focuses on leveraging intrinsic patterns and trends derived directly from date-based features, rather than historical predictions.

3.1.5. Algorithm

Random Forest Regressor: This choice of algorithm is due to its effectiveness in handling nonlinearities and its robustness in various predictive modeling scenarios. Random Forest is particularly suited for this application because it can manage the complex interactions between different time-based features without extensive tuning.

3.1.6. Feature Selection

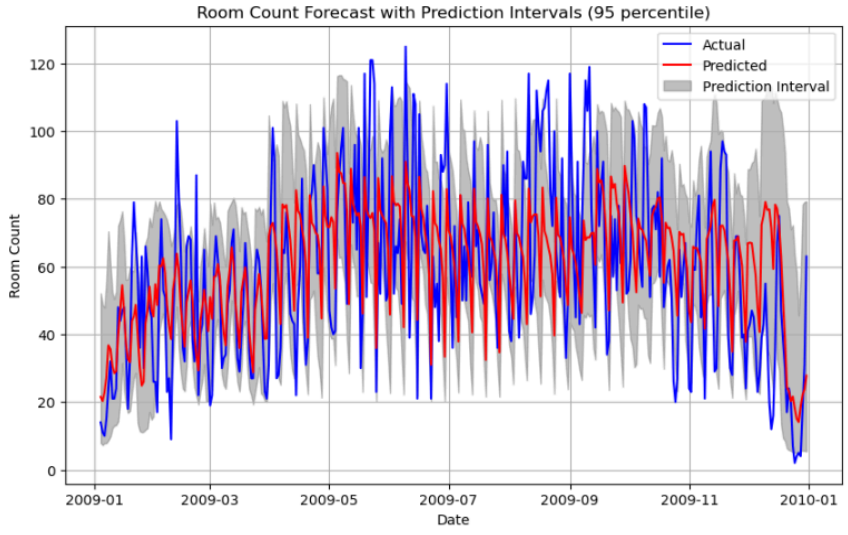
Date Features: The models utilize features derived from the date such as day of the week, month, and specific cyclical transformations (e.g., sine and cosine of day and month) to capture seasonal and weekly patterns effectively. This ensures that the models can recognize and adjust to periodic fluctuations in booking data, which are critical for accurate forecasting in the hospitality industry.

3.1.7. Overall Conclusion

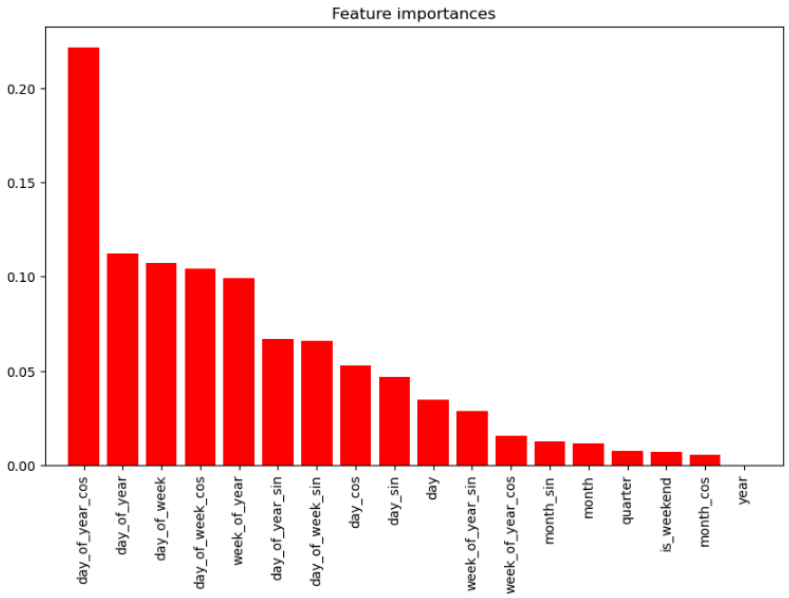
Each model is equipped to handle its designated time frame effectively, utilizing a tailored set of features that best represent the predictive signals relevant to daily, weekly, or monthly booking trends. This specialized approach enhances the models' accuracy and utility in practical applications within the hotel management sector.

3.1.8. Visualizations and Results

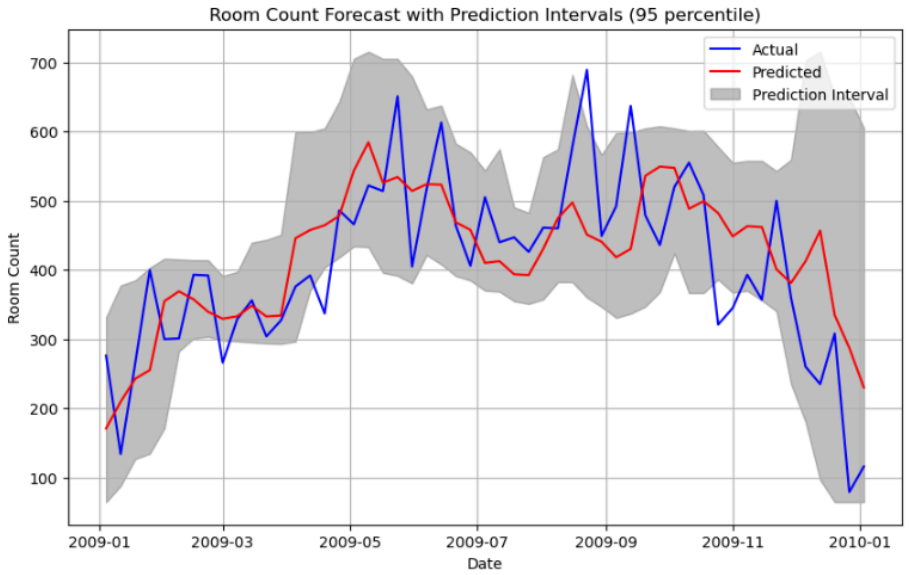
Multi-Step Day



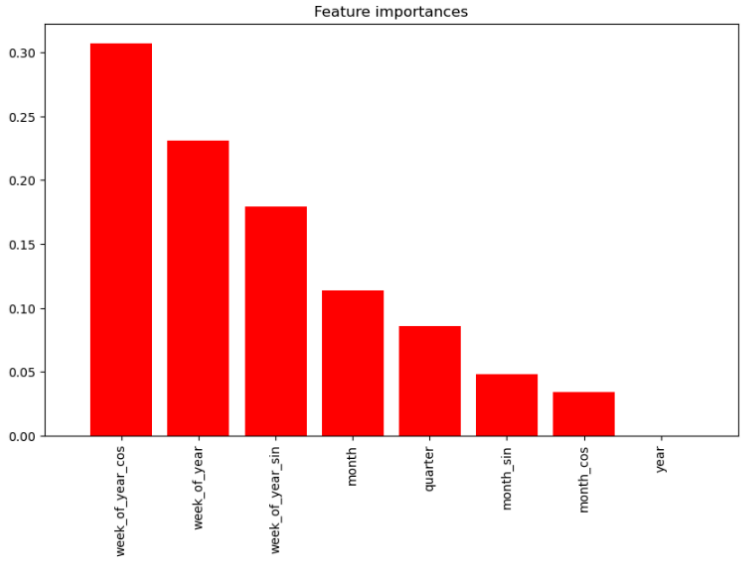

```
Percentage of actual values within predicted interval: 73.20%
Mean Squared Error: 460.63
Root Mean Squared Error: 21.46
```



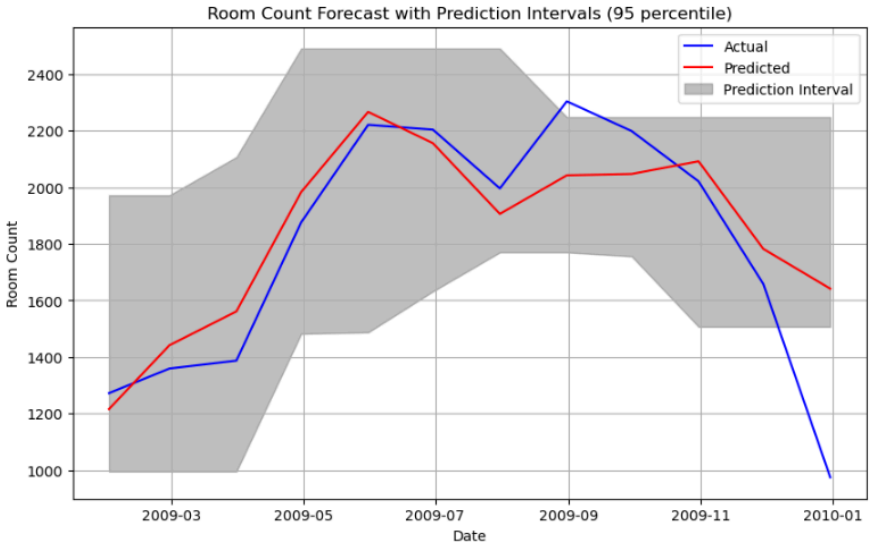
Multi-Step Week



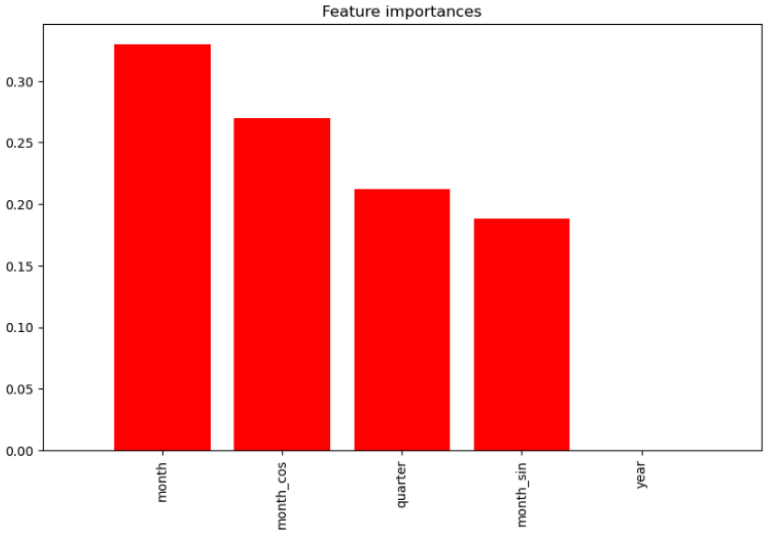

```
Percentage of actual values within predicted interval: 88.68%
Mean Squared Error: 8808.17
Root Mean Squared Error: 93.85
```



Multi-Step Month




```
Percentage of actual values within predicted interval: 83.33%
Mean Squared Error: 51802.10
Root Mean Squared Error: 227.60
```



3.2. One-Step Prediction Model for Occupied Rooms

The one-step forecasting models, specifically designed for weekly (**one_step_week**) and monthly (**one_step_month**) predictions, leverage detailed historical reservation data to forecast future room counts. Unlike multistep models that predict without utilizing past booking outcomes, one-step models integrate past booking data up to the current stay date to estimate future demands.

3.2.1. Models

- **multi_step_week**: Predicts the total number of room bookings for the following week.
- **multi_step_month**: Estimates room bookings for the upcoming month.

3.2.2. Purpose

These models are tailored to utilize historical booking patterns and actual reservation data, providing a more granular and historically informed prediction which enhances forecast accuracy for short-term planning.

3.2.3. Data Management

3.2.3.1. Data Handling Strategy

Each model processes individual reservation records, capturing detailed booking information up to the stay date.

Data is aggregated weekly or monthly, depending on the model, summarizing features to match the prediction period.

3.2.3.2. Feature Aggregation

Calculate Aggregates: Key reservation details are aggregated using specific functions designed to capture the essence of booking patterns:

- **Average and Summation Metrics:** Include average nights per stay, average lead time, average number of adults, total room bookings, and total revenue.
- **Demographic Segmentation:** Counts of reservations from specific countries, distinguishing between domestic (HR) and international guests.

3.2.3.3. Incorporation of Past Data

- **Expected Number of Room Count:** This feature, derived from the aggregation of past bookings, is critical as it provides a forecast of expected bookings based on reservations confirmed prior to the stay date. This data point is particularly valuable for enhancing the accuracy of the forecast by providing a quantifiable measure of anticipated room demand, enabling more precise adjustments in the models' output.

3.2.4. Forecasting Approach

- **One-Step Prediction Strategy:** Each model makes a single prediction for its designated time frame using a comprehensive set of features derived from past bookings. This approach allows the models to leverage recent data more effectively than multistep models, which do not use outputs from previous forecasts.

3.2.5. Feature Utilization

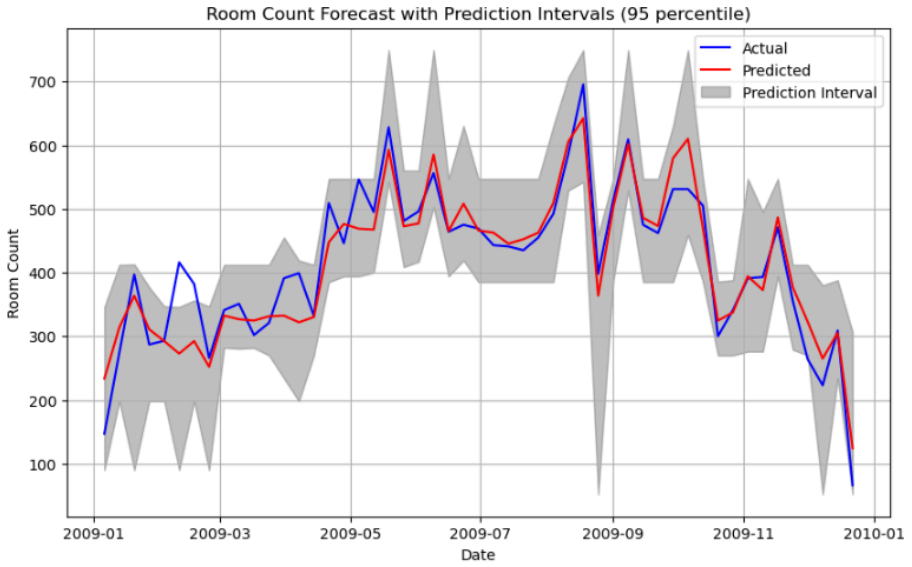
- **Date Features:** As with the multistep models, date-based features are crucial for capturing seasonal trends and cyclical booking patterns. These include day of the week, month, and special dates that could influence guest booking behavior.
- **Reservation Features:** By analyzing past reservation data, these models can identify patterns and anomalies that may not be evident from date features alone, allowing for adjustments based on recent market dynamics and operational changes.

3.2.6. Overall Conclusion

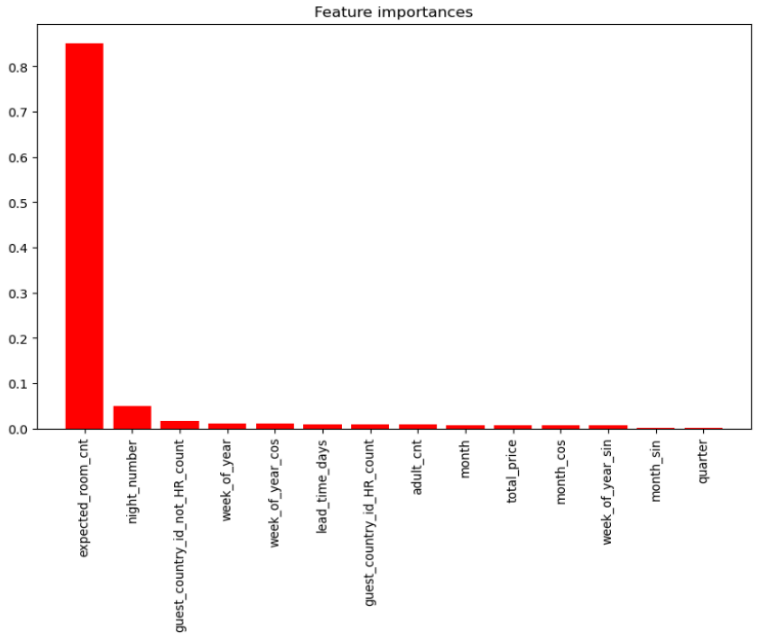
The one-step forecasting models harness the power of detailed historical data combined with sophisticated aggregation techniques to provide accurate and actionable forecasts for hotel room bookings on a weekly and monthly basis. This strategic use of data not only improves the precision of the forecasts but also enhances the hotel's ability to engage in proactive and informed decision-making.

3.2.7. Visualizations and Results

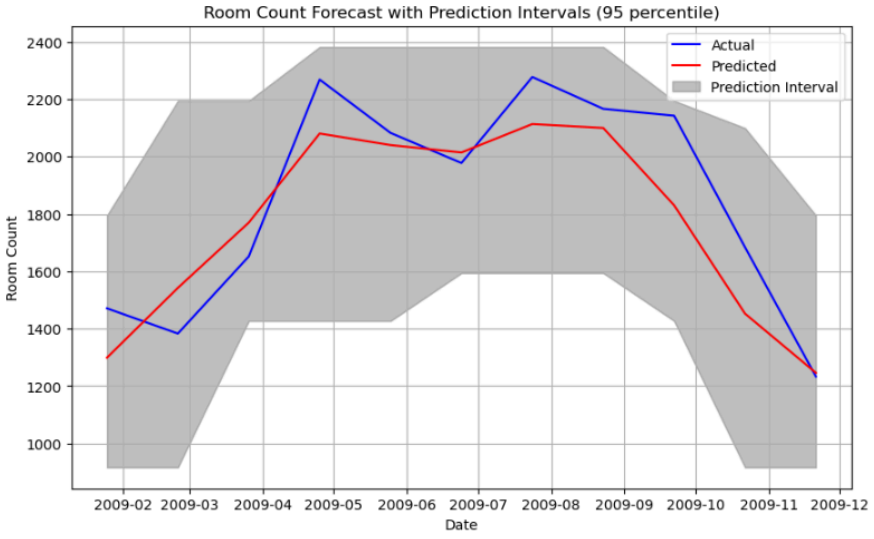
One-Step Week



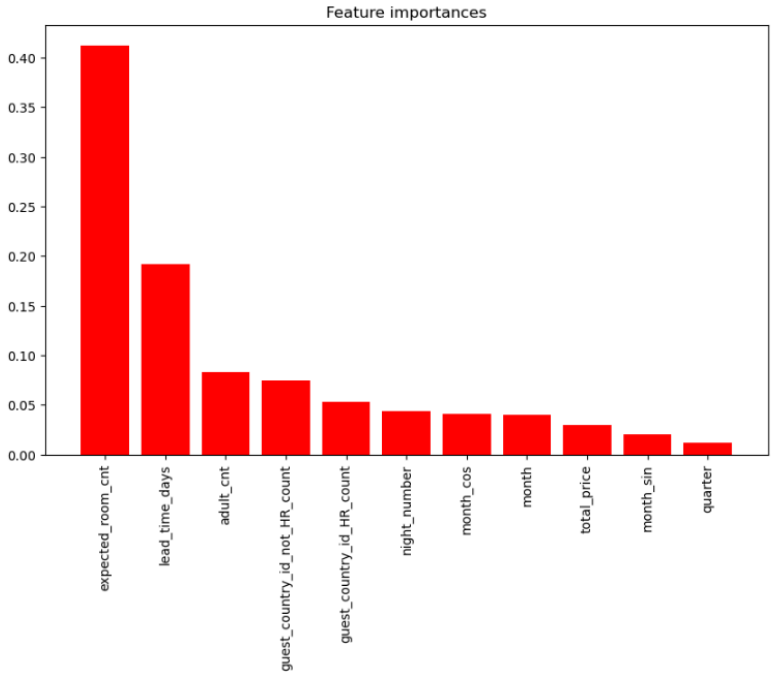

```
Percentage of actual values within predicted interval: 94.12%
Mean Squared Error: 1787.17
Root Mean Squared Error: 42.27
```



One-Step Month




```
Percentage of actual values within predicted interval: 100.00%
Mean Squared Error: 26393.74
Root Mean Squared Error: 162.46
```



3.3. Special One-Step Daily Prediction Model Utilizing Cancellation Prediction with Next Day Room Count Prediction

This innovative model is crafted to forecast hotel occupancy for the upcoming day by integrating predictions on reservation cancellations and new reservations expected within the same period. Employing a two-pronged approach, the model first utilizes historical data up to the day prior to the prediction to estimate likely reservation cancellations and incoming reservations.

3.3.1. Model Composition and Operation

The model is composed of two main components: a cancellation prediction model and a room count prediction model. The cancellation prediction aspect is handled by a Random Forest Classifier equipped with 100 estimators. This classifier assesses which existing reservations are likely to be canceled. In parallel, a Random Forest Regressor with parameters set to n_estimators=700, max_depth=6, min_samples_leaf=2, random_state=42 predicts the number of new reservations that are expected to be made for the next day. The final occupancy prediction for the next day is calculated by summing the rooms predicted by the regression model and subtracting those predicted to be canceled by the classification model. The robustness of the cancellation prediction component is demonstrated by its 92% accuracy rate on the validation set.

3.3.2. Evaluation Metrics

The overall model’s effectiveness is evaluated through a unique method: it generates a predicted range (lower bound, upper bound) for the next day's room count. A day is considered accurately predicted if the actual room count falls within this range. The final precision of the model is quantified by

calculating the percentage of days where the real occupancy count falls within these predicted bounds. The results of this model on validation set (data after March 2009 in original train set) are encapsulated through various metrics:

- **Mean Squared Error (MSE): 53**, which helps in understanding the average of the squares of the errors—the average squared difference between the estimated values and the actual value.
- **Root Mean Squared Error (RMSE): 7.3**, providing a measure of the differences between values predicted by the model and the values actually observed from the environment that is being modeled.

- **Precision Score: 81.70%**, representing the proportion of days for which the model successfully predicted the actual room count within the forecasted range.

3.3.3. Visualization and Insights

The results and effectiveness of this dual-model approach are visualized below. This model not only enhances the accuracy of daily occupancy forecasts but also provides operational insights that can significantly improve resource allocation and customer service in hotel management.

Image 1: Cancellation Prediction

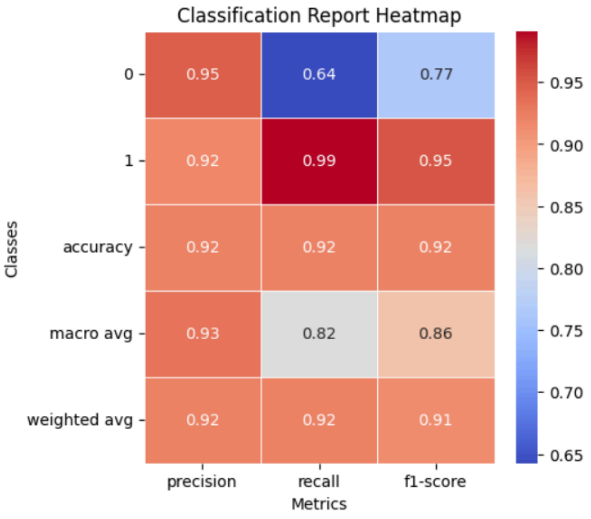


Image 2: Confusion Matrix for Cancellation Prediction

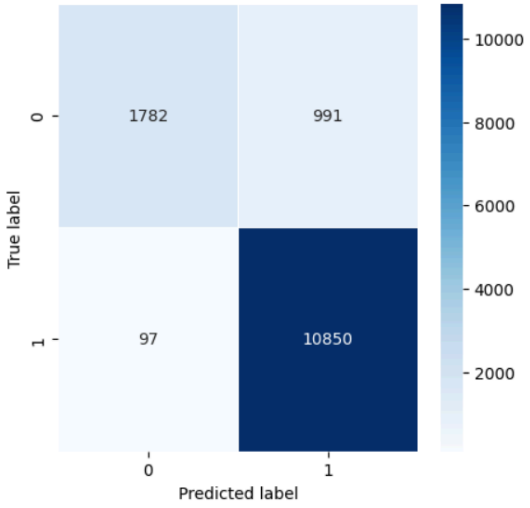


Image 3: Final Model Prediction Visualization