Data Intensive Computing

CSE 4/587 Project Phase 3

TITLE: Hotel Bookings Data Analysis

Rohit Ramichetty rohithra@buffalo.edu

Vrashi Shrivastava vrashish@buffalo.edu

1. Problem Statement

For this Project, we are going to take a Hotel Booking dataset, clean it by handling missing values and managing the format of the data, perform EDA on it to understand it better and run it through multiple models to draw intelligent information from it regarding the profits revenue, the possible bookings along with cancellations, and customer satisfaction of the hotel to ultimately predict the demand of the hotel and how can it be benefitted through data-driven

2. Model Selection:

Following five machine learning algorithms were used in this project:

1. KNN algorithm
2. Logistic Regression
3. Navies Bayes
4. Decision Tree
5. Adaboost

We have used the above-mentioned models on same data for different scenarios where each algorithm was used for particular use case which adds up to the final decision. The accuracy of all the algorithms is listed below.

Name		Accuracy
KNN Algorithm	-	72%
Logistic Regression	-	99%
Navies Bayes	-	53%
Decision Tree	-	89%
Adaboost	-	62%

After conducting tests with five different machine learning algorithms on our dataset, we found that Logistic Regression yielded the highest accuracy in predicting hotel booking waitlists. Additionally, we used each of the other algorithms to predict their respective use cases, which were then aggregated to inform our final decision on whether a particular hotel was in demand or not.

To further assess the potential profitability of each hotel, we predicted revenue, possible bookings (including cancellations), and customer satisfaction. These insights allowed us to accurately forecast demand for each hotel and make informed decisions regarding which hotels were most likely to yield successful business outcomes.

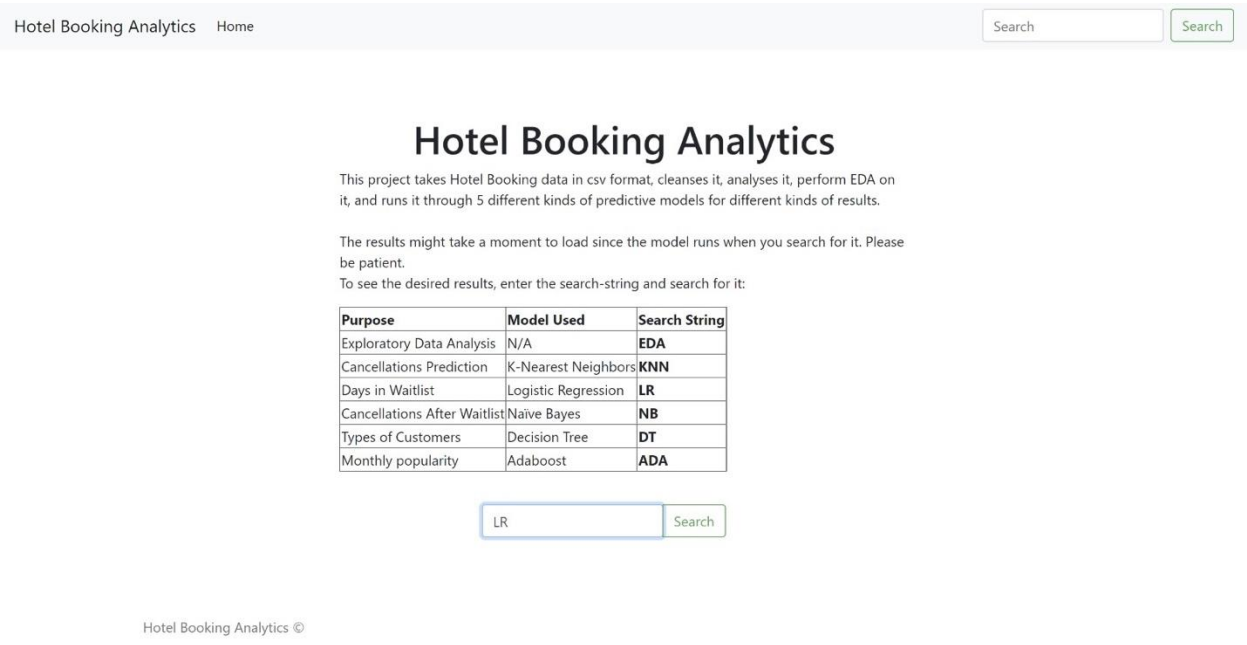
We provided recommendations to each hotel based on our findings, with the goal of optimizing their business strategies to better meet the needs of their customers and ultimately drive greater profitability. By leveraging these insights, we believe that hotels can make informed decisions and capitalize on emerging trends in the hospitality industry.

3. Web Application:

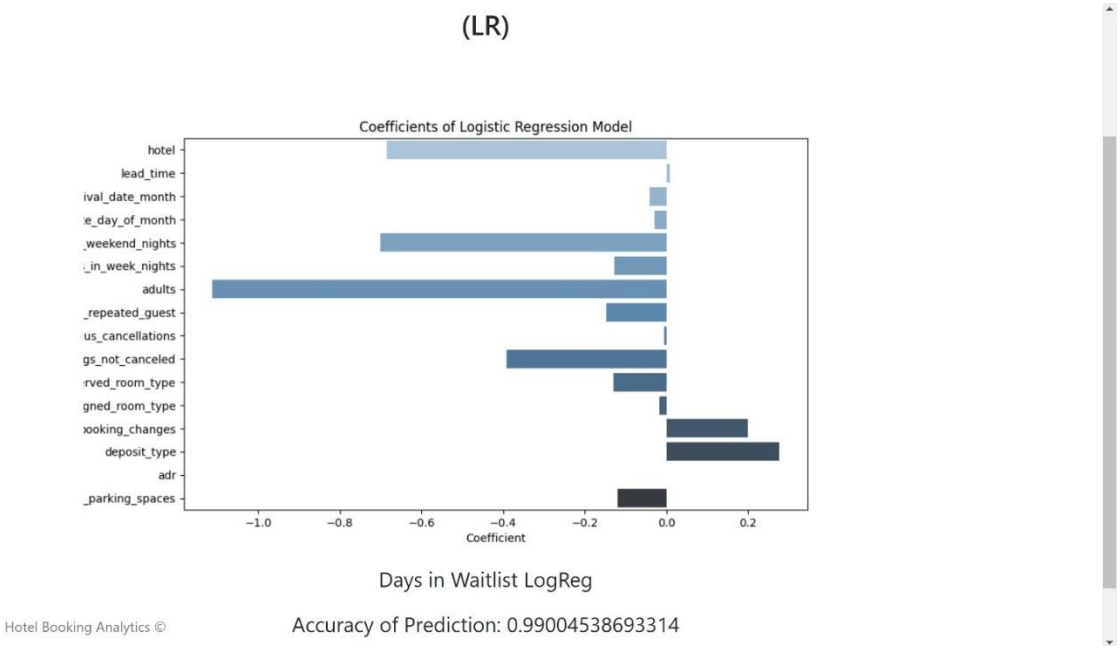
We have used the following Web Applications to develop our product:

- For Frontend, we have used HTML and CSS.
- For Backend, we have used is Flask (Python Framework).
- Machine Learning Models.

In the front-end application we must select the machine learning algorithm which we have to use and then click on search



Then we get a visualization of the model which helps us gain a deeper understanding of the waitlist prediction problem and identify the most important factors that drive waitlist demand. This information can be used by hotels to improve their revenue management strategies and allocate resources more effectively to meet customer demand.



4. How does our product help the problem:

Hotels play a critical role in providing hospitality to new visitors in a state or country. In today's competitive landscape, hotels strive to outperform their competitors by offering the best facilities and services. Our machine learning model can help hotels gain an edge in this race by providing them with valuable insights and identifying patterns that they may not have been aware of before.

By analyzing historical data on bookings, cancellations, and customer satisfaction, our model can help hotels gain a deeper understanding of what drives demand and make more informed decisions about how to allocate their resources and marketing efforts. This can allow hotels to adjust their pricing and inventory management strategies to maximize their revenue potential.

Moreover, the insights generated from our analysis can help hotels identify areas for improvement in their services and amenities. By improving customer satisfaction and loyalty, hotels can not only retain their existing customers but also attract new ones.

In addition, our model can help hotels gain a better understanding of their customers, allowing them to tailor their marketing and outreach efforts to attract more of their target audience. This can lead to stronger relationships with customers and a better reputation in the industry.

Overall, model can empower hotels to make more informed decisions, optimize their operations, and ultimately drive greater profitability.

5. Working Instruction:

The instructions are provided in the README file.

6. Extended ideas and Future scope:

For now, we have used only two types of hotels and used some patterns but extend our model using on more hotels and we could consider one potential area of expansion is to incorporate new data sources, such as weather data or social media data, to gain a more complete picture of demand patterns and customer preferences and also conducting a regional analysis to understand how demand patterns vary across different areas and regions. This could help hotels target their marketing efforts more effectively and optimize their pricing and inventory management strategies.

7. References:

- **Kaggle:** Hotel booking demand | Kaggle
- **Flask setup: Flask VSCode**
- **Flask Tutorial: Flask Python**
- **Modal:** Logistic Regression
- **Other Modals:** ML Modals