# CSE 587 DATA INTENSIVE COMPUTING

# PROJECT PHASE- 2 REPORT


### SAI TEJA SANKARNENI AND NISHANT VARMA INDUKURI – CSE 587


## PROBLEM STATEMENT

(A)

The Online Hotel Reservation system has changed the landscape of hotel booking. Customers no longer need to worry about the availability of rooms and services until they reach the front desk. They can check and plan accordingly beforehand through Online Reservation Systems. However, this has also led to a few more challenges for the hotels beyond meeting customer expectations. It has led to a significant increase in cancellations and no-shows.

Although cancellations are common occurrences and could be due to various reasons like a change of plans, weather, scheduling conflicts, etc., they have caused negative implications for the hotels. These hotels started to overbook or under book in some instances, which ultimately led to an impact on the Hotels' revenue.

So, to deal with these kinds of situations, we decided to come up with a model that can help predict whether the customer will cancel the reservation or not. The contents of the table are given below:

- Booking_ID: Unique booking code for each booking
- No_of_adults: Number of adults
- No_of_children: Number of children
- No_of_weekend_nights: Number of Saturday/Sunday nights the guests stayed
- No_of_week_nights: Number of weekday nights the guests stayed
- Type_of_meal_plan: Meal plan chosen by the guest
- Required_car_parking_space: Whether the guest require car parking or not
- Room_type_reserved: Room type selected by the customer
- Lead_time: Number of days between the date of booking and date of arrival
- Arrival_year: Year of the arrival date
- Arrival_month: Month of the arrival date
- Arrival_date: Date of the month of arrival
- Market_segment_type: Designation of market segment
- Repeated_guest: Is the guest a frequent or returning customer
- No_of_previous_cancellations: Number of previous bookings cancelled before arrival date

- No_of_previous_bookings_not_cancelled: Number of previous bookings attended by the customer.
- Avg_price_per_room: Average price of room per day of reservation
- No_of_special_requests: Number of special requests made by the customer
- Booking_status: Booking indication whether it is cancelled or not

(B)

This project has the potential to revolutionize how hotels manage their booking, increase revenue, and enhance customer satisfaction. With the help of this model, we can use historical data on hotel bookings to optimize prices, predict future demand, and suggest choices and services to customers. Through this, hotels could be well-functioning and aim for profits.

## DATA SOURCES

The data file is also attached in the folder as Hotel_reservations_data.csv

And has been taken from Kaggle. It consists of 36275 rows and 19 columns.

[https://www.kaggle.com/datasets/ahsan81/hotel-reservations-classification-dataset?datasetId=2783627&sortBy=voteCount](https://www.kaggle.com/datasets/ahsan81/hotel-reservations-classification-dataset?datasetId=2783627&sortBy=voteCount)

## MODEL IMPLEMENTATION:

After Performing Cleaning and Exploratory Data Analysis on our dataset we have performed 5 Machine Learning Models in Phase-2 of our project based on various evaluation metrics observed while performing the Exploratory Data Analysis. We have Performed 4 different algorithms for Classification (To predict Cancellation Status) and a linear regression model to predict the average price per room for a given booking.

| Model | Accuracy (%) | Precision Score (%) |
|---|---|---|
| Logistic Regression | 81.52 | 92.31 |
| Naïve Bayes | 36.90 | 16.26 |
| Support Vector Machines | 78.42 | 97.42 |
| Decision Tree Classifier | 79.55 | 82.88 |

For, **Linear Regression Model** used to predict the average price per room of a booking. The mean squared error is 0.0025 and the root mean squared error is 0.050. We have also derived the Mean absolute error; it is also returned as 0.037 and the R2 score is 0.4609.

## MODEL SELECTION:

We are selecting Logistic Regression Algorithm out of all the Machine Learning models we have performed in Phase-2 to predict the Customer Cancellation trend and implement it in our front-end application.

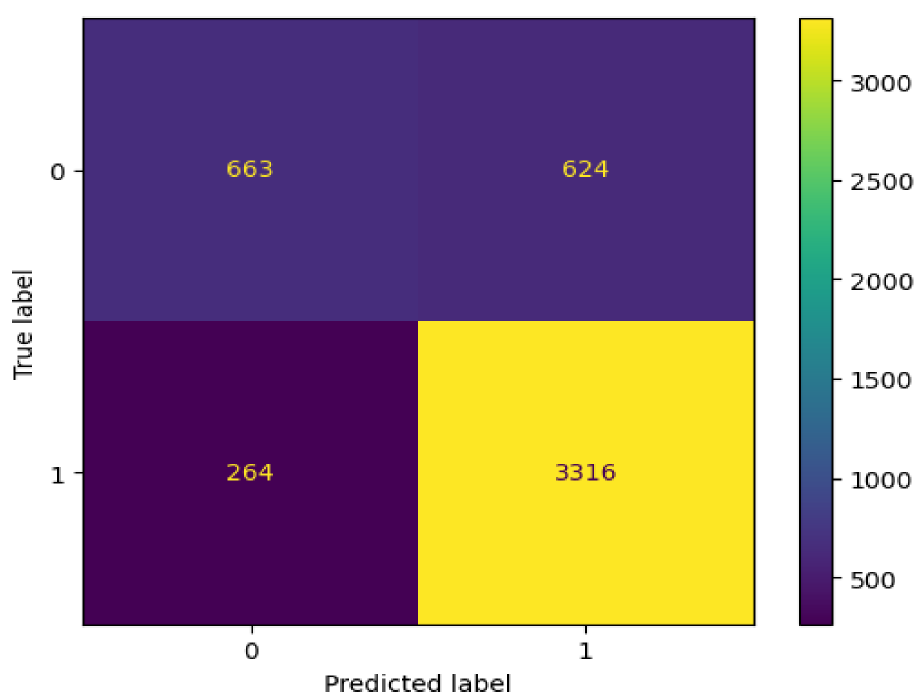
We have chosen Logistic Regression due to its ease of implementation and also we have achieved 81.52 % Accuracy (Highest compared to all other models we have performed).

## MODEL EVALUATION METRICS:

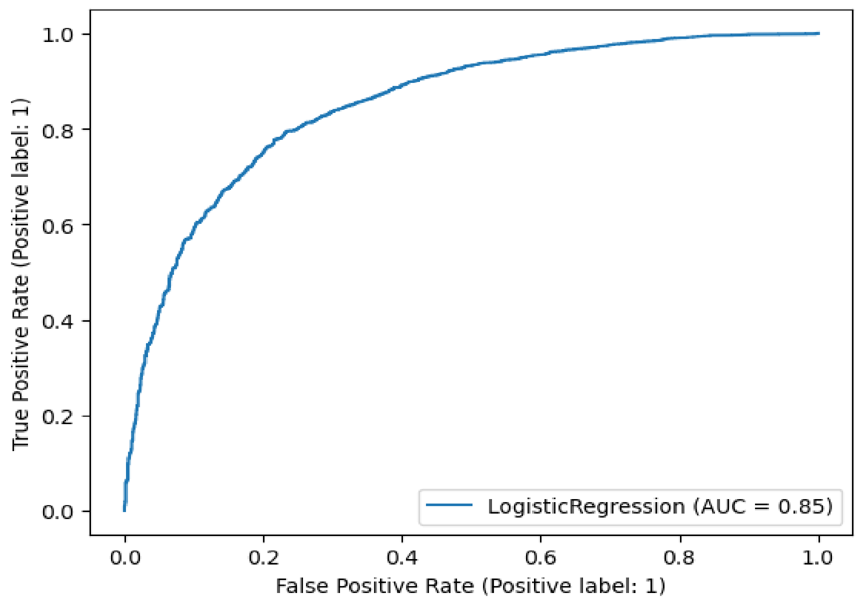
**Various Scores:**

```
Model: Logistic Regression
Different Metrics of the Model on Testing Data:
Accuracy: 81.75467433737415%
Precision Score: 92.62569832402234%
f1 Score: 88.19148936170212%
Recall_Score: 84.16243654822335%
```

**Confusion Matrix:**



**ROC Curve:**




## WEB APPLICATION

Tech Stack we have used to develop our web application are:

- Frontend: HTML, CSS
- Backend: Flask (Python Framework)
- Plotting Graphs:
- ML Model: XG Boost

## INSTRUCTIONS TO RUN:

**STEP 1:**

We have included the Web application and its working directory. Open the directory with all the files of the project. Below is a screen shot of these

**STEP 2:**

Run the app.py file and the flask_server.py in the local host to start the flask API

**STEP 3:**

Open the following URL in your web browser http://127.0.0.1:5000/ to run the web application.

The following is our initial UI as we enter the Web Application

We then need to fill out the required details to obtain the predictions about whether the customer would be cancelling the booking or else not. Which is shown below with an example

# Customer Reservation Cancellation Prediction

**Total no.of Adults**

**Total no.of children**

**Total nights over the weekend**

**Total nights over the week**

**Type of Meal Plan**

--Select--

**Requirement for Car Parking**

--Select--

**Type of Room Reserved**

--Select--

**Lead Time**

**Type of Market Segment**

--Select--

**Previous Customer? (Repeated guest)**

--Select--

**Total # of Previous Cancelled Bookings**

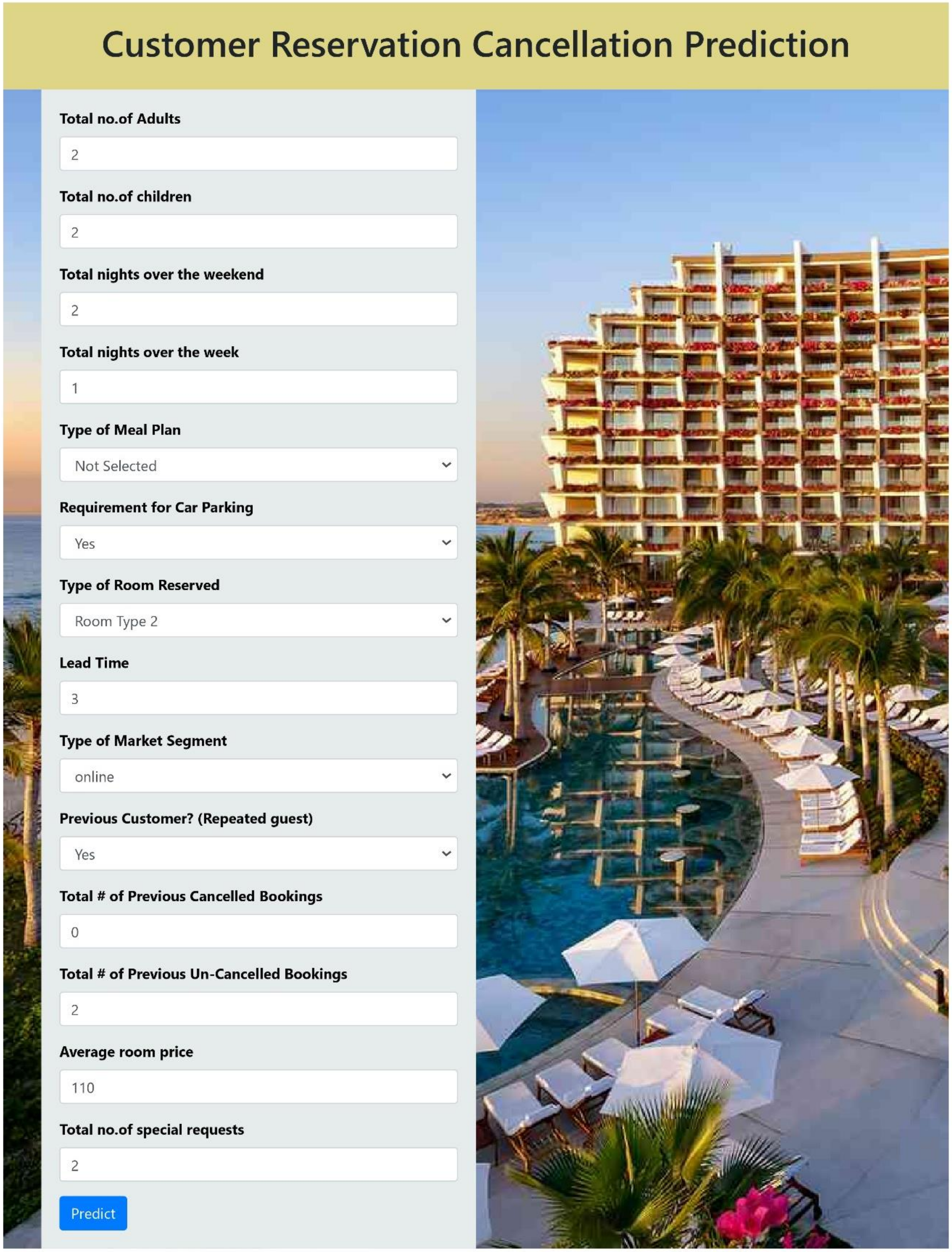
**Total # of Previous Un-Cancelled Bookings**

**Average room price**

**Total no.of special requests**

Predict [Click here to visit the Dashboard]

**EXAMPLE:**


Customer Reservation Cancellation Prediction
Total no.of Adults
2
Total no.of children
2
Total nights over the weekend
2
Total nights over the week
1
Type of Meal Plan
Not Selected
Requirement for Car Parking
Yes
Type of Room Reserved
Room Type 2
Lead Time
3
Type of Market Segment
online
Previous Customer? (Repeated guest)
Yes
Total # of Previous Cancelled Bookings
0
Total # of Previous Un-Cancelled Bookings
2
Average room price
110
Total no.of special requests
2
Predict

**The output we get to the above example is**

5/10/23, 10:19 PM Customer Booking status Prediction

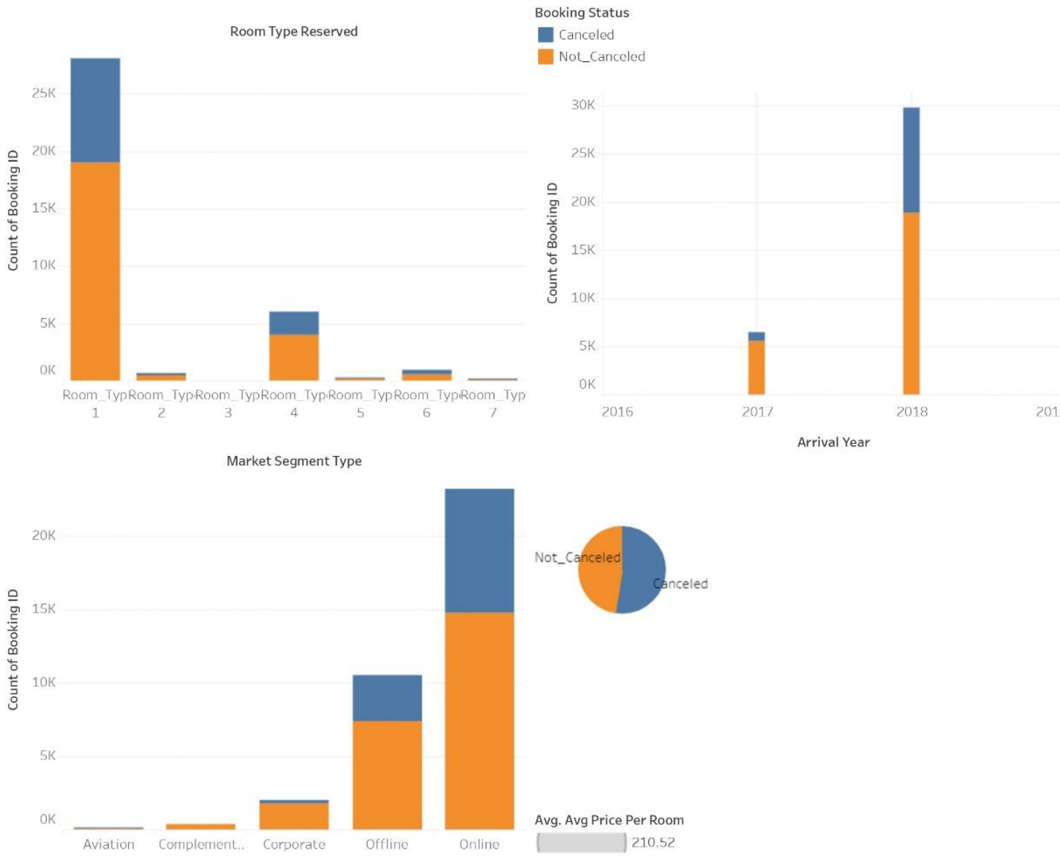
# Prediction: 1 with Probability: 0.9978960063486468

Relax!! The Booking will not be Cancelled


**VISUALISATION SHOWN IN THE UI**

Click here to visit the Dashboard

This button is present beside the predict button which takes us to the visualisation of the predictions made and the user can look at the correctness of the model.


Room Type Reserved
Count of Booking ID
25K
20K
15K
10K
5K
0K
Room_Typ 1
Room_Typ 2
Room_Typ 3
Room_Typ 4
Room_Typ 5
Room_Typ 6
Room_Typ 7
Booking Status
Canceled
Not_Canceled
30K
25K
20K
15K
10K
5K
0K
2016
2017
2018
Arrival Year
Market Segment Type
Count of Booking ID
20K
15K
10K
5K
0K
Aviation
Complement..
Corporate
Offline
Online
Not_Canceled
Canceled
Avg. Avg Price Per Room
210.52