**High Level Design (HLD)**

Hotel Churn Rate Prediction using Machine Learning

Revision Number: 1.0

Last date of Revision: 31-07-2022

**Document Version Control**

|  |  |  |  |
| --- | --- | --- | --- |
| Date Issued | Version | Description | Author |
| 31-07-2022 | 1 | Initial HLD | Aravind |

ABSTRACT……………………………………………………………………………………… 1

1. INTRODUCTION…………………………………………………………………………. 1
   1. Why this High-Level Design Document? ……………………………….. 2
   2. Scope………………………………………………………………………………………2
   3. Definitions……………………………………………………………………………… 3
2. GENERAL DESCRIPTION………………………………………………………………..3

Abstract

Reservation is to arrange your Room before going there. The guests indicate the room number, type and the time he/she came in advance and hotels try to adjust their preparations and the needs to be provided for the guests accordingly. It is also critical for hotels to be informed about reservation in advance in terms of hotel expenses. Booking Cancellation has significant effect on revenue of the Hotel. To solve the cancellation problem and its effect, hotel management builds a machine-learning model, by using data science tools and techniques this model aims to predicts hotel booking cancellation by getting user inputs through web application. By implementing various Supervised Classification Machine learning algorithms like, Logistic Regression, Naïve Bayes, Random Forest, Gradient Boosting, KNN, XGBClassifier, Cat Boosting, Support Vector Classifier, Ada Boosting and we choose one model which performs better.

1. **INTRODUCTION**

* Booking cancellations have a substantial impact in demand management decisions in the hospitality industry.
* Cancellations limit the production of accurate forecasts, a critical tool in terms of revenue management performance.
* To circumvent the problems caused by booking cancellations, hotels implement rigid cancellation policies and overbooking strategies, which can also have a negative influence on revenue and reputation.
* Using data sets from four resort hotels and addressing booking cancellation prediction as a classification problem in the scope of data science.
* Results allow hotel managers to accurately predict net demand and build better forecasts, improve cancellation policies, define better overbooking tactics thus improve on inventory allocations.
* This Model can Predict whether the customer is going to Cancel the booking or not. While getting input from the user through web application.
  1. **Why High-Level Design Document?**

The purpose of this High-Level Design (HLD) Document is to add the necessary detail to the current project description to represent a suitable model for coding.

This document is also intended to help detect contradictions prior to coding, and can be used as a reference manual for how the modules interact at a high level.

The HLD will:

● Present all of the design aspects and define them in detail

● Describe the user interface being implemented

● Describe the hardware and software interfaces

● Describe the performance requirements

● Include design features and the architecture of the project

● List and describe the non-functional attributes like:

○ Security

○ Reliability

○ Maintainability

○ Portability

○ Reusability

○ Application compatibility

○ Resource utilization

○ Serviceability.

* 1. **Scope**

The HLD documentation presents the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), and technology architecture. The HLD uses non-technical to mildly-technical terms which should be understandable to the administrators of the system.

* 1. **Definitions**

|  |  |
| --- | --- |
| **Terms** | **Description** |
| Database | Collection of all the information monitored by this system |
| IDE | Integrated Development Environment |

1. **GENERAL DESCRIPTION**
   1. **Product Perspective**

The Hotel Industry has multimillion-dollar industry, and they are facing issue in Resource management and loss in Revenue because people cancel their booking before arrival date.

This is Booking Prediction Solution System is a Machine learning based system. Which predicts whether customer will cancel the booking or not, So Results allow hotel managers to accurately predict net demand and build better forecasts, improve cancellation policies, define better overbooking tactics thus improve on inventory allocations.

* 1. **Problem Statement**

In this Project we will work with Hotel Management, to develop a machine learning algorithm that predicts if a particular booking gets cancelled or not

Using data sets from resort hotels and City hotels and addressing booking cancellation prediction as a classification problem in the scope of data science.

This dataset contains 119390 observations for a City Hotel and a Resort Hotel. Each observation represents a hotel booking between the 1st of July 2015 and 31st of August 2017, including booking that effectively arrived and booking that were cancelled.

The Dataset Consists of 36 Features including Email, Name and Credit card number.

* 1. **Proposed Solution**

The main Objective of the Project is to predict If the booking will get cancelled before the arrival date. hotel has to guarantee rooms to customers who honour their bookings but, at the same time, has to bear with the opportunity cost of vacant capacity when a customer cancels a booking or does not show up, to solve this problem we are implementing Machine learning model

This can also be implemented on hotel booking website, based on the data if the customer is about to cancel the booking, Hotel can give some cancellation fees for such Bookings, so that it can compensate some part of the loss.

This model, by predicting the probability of each booking to be cancelled, could help produce better forecasts and reduce uncertainty in management decisions. This is very important in the context of revenue management, for inventory allocation and pricing decisions.

* 1. **Technical Requirements**

This document addresses the requirements for detecting prediction possibility of a bookings will get cancelled based on user input.

* 1. **Data Requirements**

Since this is hotel real data, all data elements pertaining hotel or costumer identification were deleted.

Four Columns, 'name', 'email', 'phone number' and 'credit\_card' have been artificially created and added to the dataset.

We have dataset from Real City Hotels and Resorts for a period of Two years

With 36 features.

**hotel:** The datasets contain the booking information of two hotel. One of the hotels is a resort hotel and the other is a city

**is\_canceled:** Value indicating if the booking was canceled (1) or not (0).

**lead\_time:** Number of days that elapsed between the entering date of the booking into the PMS and the arrival date.

**arrival\_date\_year:** Year of arrival date

**arrival\_date\_month:** Month of arrival date with 12 categories: “January” to “December”

**arrival\_date\_week\_number:** Week number of the arrival date.

**arrival\_date\_day\_of\_month:** Day of the month of the arrival date

**stays\_in\_weekend\_nights**: Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel.

**stays\_in\_week\_nights**: Number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel BO and BL/Calculated by counting the number of week nights.

**adults:** Number of adults**, children:** Number of children**, babies:** Number of babies

**meal: BB =** Bead and Breakfast, which is the most common meal basis encountered in hotels. Breakfast is included in the room rate**.**

**FB =** Full Board and means that all three meals are included – breakfast, lunch, and dinner.

**HB (Half Board) =** both breakfast and evening meal are included in your room rate.

**SC =**Self Catering. **Undefined:** Not defined

**country:** Country of origin

**market\_segment:** Market segment designation. In categories, the term **“TA”** means “Travel Agents”and **“TO”** means **“**Tour Operators”

**distribution\_channel:** Booking distribution channel. The term **“TA”** means“Travel Agents”and **“TO”** means“Tour Operators”

**is\_repeated\_guest :** Value indicating if the booking name was from a repeated guest (1) or not (0)

**previous\_cancellations:** Number of previous bookings that were cancelled by the customer prior to the current booking

**previous\_bookings\_not\_canceled:** Number of previous bookings not cancelled by the customer prior to the current booking

**reserved\_room\_type**: Code of room type reserved. Code is presented instead of designation for anonymity reasons

**assigned\_room\_type:** Code for the type of room assigned to the booking. Sometimes the assigned room type differs from the reserved room type due to hotel operation reasons.

**booking\_changes:** Number of changes/amendments made to the booking from the moment the booking was entered on the PMS until the moment of check-in or cancellation

**deposit\_type: No Deposit** – no deposit was made; **Non Refund** – a deposit was made in the value of the total stay cost; **Refundable** – a deposit was made with a value under the total cost of stay.

**agent:** ID of the travel agency that made the booking

**company**: ID of the company/entity that made the booking

**days\_in\_waiting\_list:** Number of days the booking was in the waiting list before it was confirmed to the customer

**customer\_type: Group** – when the booking is associated to a group; **Transient –** when the booking is not part of a group or contract, and is not associated to other transient booking; **Transient-party –** when the booking is transient, but is associated to at least other transient booking

**adr:** Average Daily Rate (Calculated by dividing the sum of all lodging transactions by the total number of staying nights)

**required\_car\_parking\_spaces:** Number of car parking spaces required by the customer

**total\_of\_special\_requests:** Number of special requests made by the customer (e.g. twin bed or high floor)

**reservation\_status: Check-Out –** customer has checked in but already departed; **No-Show –** customer did not check-in and did inform the hotel of the reason why

**reservation\_status\_date:** Date at which the last status was set**.**

* 1. **Tools Used**

Python programming language and modules such as Numpy, Pandas, Scikit-learn, Category Encoders, Streamlit, Dockers are used to build the whole project.



* VS Code is used as IDE for python
* For Visualization Matplotlib, Seaborn, Plotly are used.
* Front end and App are used for streamlit framework.
  1. **Constraints**
  2. **Assumptions**