# Cover and Title Page

# SUPERVISOR'S RECOMMENDATION

# CERTIFICATE OF APPROVAL

# ACKNOWLEDGEMENT

Our team would like to extend our immense gratitude to our esteemed supervisor Er. Dhiraj Kumar Jha, who showed us exceptional support and guidance throughout the development of this project. We want to convey our heartfelt gratitude to teachers as well whose valuable feedback and substantial contributions have been extremely helpful in the success of this project.

# ABSTRACT

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF ABBREVIATIONS

# CHAPTER 1: INTRODUCTION

## 1.1 Introduction

In the dynamic and fast-paced hospitality sector, successfully managing hotel reservations is critical for hoteliers. The uncertainty connected with booking cancellations is one of the issues that hotels encounter.

Hotel Booking Cancellation Prediction System is a web-based application that helps in mitigating the risks associated with reservations using machine learning algorithms. This web system provides two user bases, customers and employees. Customers can make their reservations for hotel rooms through the app while employees can view the reservations and make predictions on whether the customer is likely to cancel or not. This system attempts to give hoteliers useful information to improve decision-making, and optimize revenue management by utilizing tree-based machine learning algorithms.

## 1.2 Problem Statement

Hoteliers in the hospitality sector have the difficult challenge of efficiently handling reservations in the context of unpredictable cancellations. These cancellations have a real influence on revenue management and customer satisfaction in addition to interfering with operational planning. One major obstacle that leads to operational inefficiencies and inefficient use of resources is the lack of a systematic method for forecasting reservation cancellations. Moreover, a flexible solution is required due to the dynamic nature of the hospitality industry, which is impacted by various factors.

The necessity for an extensive booking cancellation prediction is highlighted by the fact that many hotels continue to use outdated approaches and lack the resources to fully utilize data analytics and machine learning for predictive insights into booking cancellations. Machine learning plays a crucial role in predicting hotel booking cancellations by analysing patterns, trends, and various factors that contribute to customer decisions.

In conclusion, the development and implementation of the Hotel Booking Cancellation Prediction System addresses the complex challenges faced by the hospitality industry,

providing a strategic and proactive solution for effectively managing and mitigating the impact of reservation cancellations.

## 1.3 Objectives

The primary objectives of the project are:

- To implement random forest algorithm and several optimization strategies for model development.
- To use machine learning for forecasting cancellations of reservations.
- To create a web application and incorporate a model that reliably classifies reservations.

## 1.4 Scope and Limitations

The system takes data from customer's booking requests such as meal plan, number of adults, number of children, meal plan, lead time, etc. as inputs to the machine learning model which then predicts the cancellation of bookings based on these provided data. The model uses Random Forest algorithm to predict booking cancellations.

The limitations of the system are as follows:

- The dataset used for this project only contains the data from branches of a single hotel, not multiple hotels.
- There is no limit to the number of reservations done by a user.

## 1.5 Development Methodology

Plan-driven incremental development approach is used for developing this project. Incremental development model helps in building software systems components-by-components in which the final requirements are clear from the start. It involves developing initial version, providing feedback and changing software through different versions until the required system is developed. Hence, the incremental methodology develops the system as a series of increments or versions, with each version adding functionality to the previous version.

The following figure shows the different phases of incremental development:

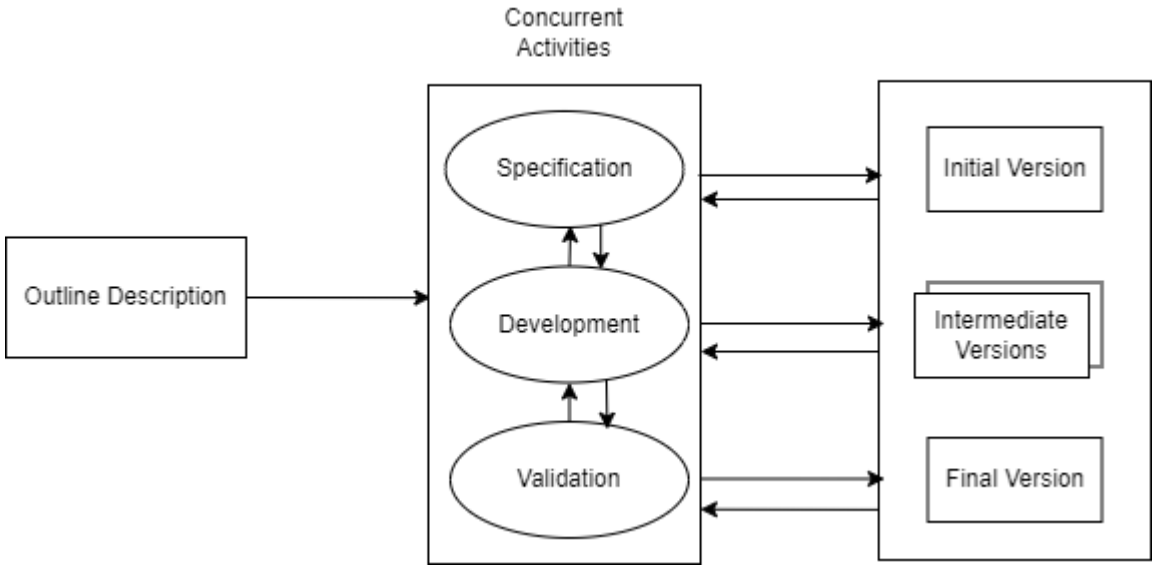


*Figure 1.1: Incremental Delivery Model*

## 1.6 Report Organization

The report consists of six chapters which are organized as follows:

**Chapter 1: Introduction -** This chapter includes a thorough project introduction, problem statements that correspond with the project, the project's objectives, its scope and limitations, and the methods employed to build the system.

**Chapter 2: Background Study and Literature Review** - This chapter offers a review of earlier research on the issue. It covers various study related to data processing and machine learning techniques in order to implement the project.

**Chapter 3: System Analysis -** This chapter explores both the system's functional and non-functional requirements. It also includes a feasibility study conducted to examine the system's operation and the system's work breakdown structure.

**Chapter 4: System Design -** This chapter includes detailed design of the system along with algorithm details.

**Chapter 5: Implementation and Testing** - This chapter covers the hardware tools, dependencies, and software tools needed to complete the project. Several project implementation steps are outlined in this chapter along with various test cases.

**Chapter 6: Conclusion and Future Recommendations** - This chapter analyses the outcome of the model that was put into practice. It also offers additional tasks that can be performed to improve the project.

# CHAPTER 2: BACKGROUND STUDY AND LITERATURE REVIEW

## 2.1 Background Study

Classification Algorithms like the Random Forest Classifier are designed to assign predefined labels to input data based on patterns observed during the training phase. A classifier's purpose is to learn a mapping between input features and the appropriate output labels in order to generate accurate predictions on previously unseen data. Random Forest Classifier is an ensemble learning approach based on decision trees. Random Forest divides the training data into multiple subsets, with each subset utilized to train a decision tree. The number of people, arrival date, special services, car parking, and other factors all play a role in the outcome of a hotel reservation. The algorithm divides the outcome into two categories: likely to cancel and unlikely to cancel. Multiple performance indicators like Accuracy, Precision, Recall, F1-Score, and the AUC-ROC Curve will be examined to test the model's performance.

## 2.2 Literature Review

Cancellations are a crucial component of hotel revenue management due to their effect on reservations for rooms but only little is understood by hoteliers about why consumers cancel their bookings or how to prevent it. Because of the effect that cancellations have on hotel chains, hotels come up with various tactics for the express purpose of minimizing cancellations, which in turn has an influence on hotel income and reputation. These factors make it essential for hotel management to get early notice of cancellations. Demand forecasting and revenue management have a significant correlation and this hole may be filled by utilizing machine learning and various algorithms to identify those who are likely to cancel.

In order to develop domain expertise for the project, current research papers that have been published in the specialized fields of machine learning and hospitality were thoroughly

analysed. These publications' insights, which aided in supplying knowledge about the field, are mentioned below.

A Study by Eleazar C. Sanchez, Agustin J. Sanchez-Medina, Monica Pellejero on 'Identifying Critical Hotel Cancellations using Artificial Intelligence' published in 2020, conducted a comprehensive analysis on the most important factors that affect hotel booking cancellations. The research focused on predicting hotel booking cancellations made close to the time of service. The main purpose of this research was to help hoteliers improve their strategies fox maximizing revenue while minimizing risks and losses. Multiple Artificial Intelligence and Machine Learning models were applied to Personal Name Records (PNR) data which produced great results. The dataset was provided by a 4-star hotel located in Gran Canaria, Spain containing more than 10,000 booking records between 2016 and 2018. A total of 13 columns were used as independent variables and the state of booking (cancelled or not cancelled) was used as the dependent variable. The independent variables included nationality, number of nights, hotel type, previous cancellations, number of adults, number of children, and more. Models were developed on R statistical software using the following packages: C5.0, Support Vector Machine (SVM), Artificial Neural Networks (ANN) and Gradient Boosting Machine (GBM), among which GBM yielded the best results with a 73% accuracy in the worst case and an 80.5% accuracy in the best case. On the other hand, ANN produced the least favourable output with an accuracy of 60% and 69% in the best and the worst cases respectively [1].

Another research conducted by Yaqi Lin on July 2023, dealt with identifying the factors that have the greatest impact on hotel booking cancellations through Exploratory Data Analysis (EDA) visualization. Due to external uncertainties such as flight cancellations, itinerary changes, conference cancellations, many customers choose to cancel their orders after booking a room [2]. Machine Learning algorithms were applied to guess whether a customer will cancel their bookings. The algorithms used in this research were decision trees, logistic regression, and random forest. 50,000 records were randomly selected from 1,19,391 data for analysis. It was found that city hotels have a cancellation rate that exceeds that of resorts by about 12%. Moreover, July and August are the peak period for customer orders, and it is also the maximum number of cancelled orders. Other factors, such as market segment, payment deposit, advanced booking were also found to have major

impacts on booking cancellations. Among the three algorithms used, logistic regression was the weakest and random forest showed the highest accuracy results.

A paper by Rachel Mytnik, published on June 25, 2021 aimed to find the best classification model for predicting hotel booking cancellations and find the most relevant explaining variables for customer cancellations [3]. This research paints a clearer picture on hotel booking demands. The data set consists of 117,864 records from a resort hotel in Portugal. Several classification models were used in this research, including XGBoost, Random Forest, Linear SVM, and Logistic Regression amongst which XGBoost showed the highest level of accuracy of 99.7%. Linear SVM showed the lowest level of accuracy of 78.65% which was attributed to its linearity. Lastly, the variables that had the highest impact on the model were found to be deposit type followed by required parking space, previous cancellations, and so on.

The popularity of online reviews is causing a huge impact on consumers' purchase intentions for goods and services [4]. However, and hidden by the anonymity of the Internet, fraudsters can try to manipulate other consumers by posting fake reviews. Research carried out by M.R. Martinez-Torres, and S.L. Toral used a content analysis approach based on a set of unique attributes and sentiment orientation of reviews to identify deceptive reviews in the hospitality sector. The main contributions of the paper are a set of polarity-oriented unique attributes able to distinguish deceptive and non-deceptive reviews and the main topics associated to deceptive and non-deceptive reviews. The training data set was processed using various machine learning algorithms including KNN, Logistic Regression, SVM , Random Forest, Gradient Boosting, and MLP. Analyzing the data, it was found that deceptive positive reviews often emphasize the location of the hotel while positive non-deceptive reviews are focused on the characteristics of the city. The topics of negative deceptive opinions were focused on complaints about the hotel environment, room environment, etc. and in the case of non-deceptive negative opinions, the complaints are related to long waits, staff behaviour, smell, etc.

Cancellation prediction models are advantageous because they classify the cancellation outcome of each booking and allow an understanding of how each feature influences cancellations, that is, an understanding of cancellation drivers [5]. This study on 'Big data in Hotel Revenue Management' by Nuno, Almeida and Luis determined that by identifying

the features that are most important in predicting the outcome of a booking, the cancellation drivers can be narrowed down.

# CHAPTER 3: SYSTEM ANALYSIS

## 3.1 System Analysis

### 3.1.1 Requirement Analysis

An essential stage in the creation of any system is requirement analysis. It involves collecting, recording, and evaluating the requirements and expectations that help in defining the project's specifications. Both functional and non-functional requirements analysis are included in requirement analysis. The precise behaviours and functionalities that a system must have are referred to as the functional requirements whereas the characteristics that the system must have but do not directly relate to particular functionalities or behaviours are the non-functional requirements.

#### i. Functional Requirements

Functional requirements are an essential part of the development of software and systems. They define the exact functionalities, features, and capabilities that a system must have in order meet its users' needs. These criteria guide the development process by serving as a road map for designers, developers, and testers.

The functional requirements of the system are as follows:

- To allow customers and employees to register and login to the system.
- To allow customers to request bookings.
- To allow customers to cancel bookings.
- To allow employees to view customer bookings.
- To allow employee to view cancellation predictions.
- To allow employee to update booking status.

A use case diagram is a type of Unified Modelling Language (UML) diagram that visually shows the interactions between several actors and a system. It gives a high-level overview of the system functionality and how users and external entities interact with it. During the

early stages of software development, use case diagrams are frequently used to capture and explain the system's needs.

The following use case diagram for the Hotel Booking Cancellation Prediction System shows the interaction between actors (customer and employee) with the system depicting the functional requirements:

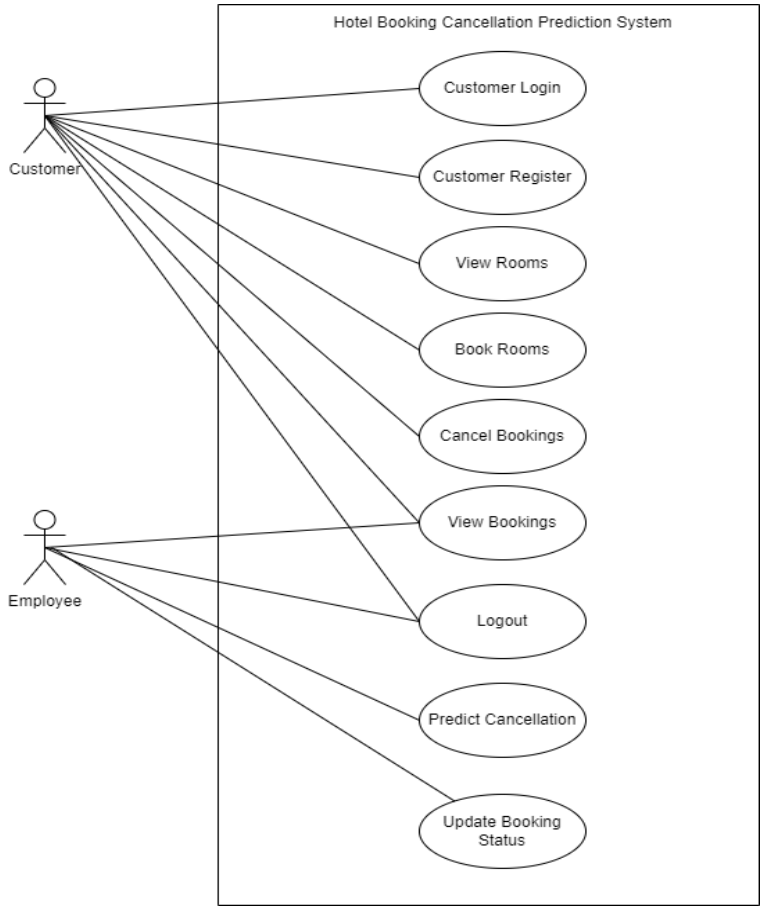


*Figure 3.1: Use Case Diagram of Hotel Booking Cancellation System*

**Use Case Description:**

Here are some of the use case descriptions of the project:

*Table 3.1: Use Case Description for Customer Registration*

| | |
|---|---|
| **Use Case ID** | UC-01 |
| **Use Case Name** | Customer Registration |
| **Primary Actor** | Customer |
| **Secondary Actor** | |
| **Description** | Registers customer to the system. |
| **Pre-Condition** | |
| **Success Scenario** | Customer can login to the system.<br>Customer data is stored in the database. |
| **Failure Scenario** | Customer is redirected to register again. |

*Table 3.2: Use Case Description for Customer Login*

| | |
|---|---|
| **Use Case ID** | UC-02 |
| **Use Case Name** | Customer Login |
| **Primary Actor** | Customer |
| **Secondary Actor** | |
| **Description** | Logs customer in to the system. |
| **Pre-Condition** | Customer must be registered to the system. |
| **Success Scenario** | Customer is redirected to the home page. |
| **Failure Scenario** | Customer is redirected to login again. |

*Table 3.3: Use Case Description for Booking Request*

| | |
|---|---|
| **Use Case ID** | UC-03 |
| **Use Case Name** | Booking Request |
| **Primary Actor** | Customer |
| **Secondary Actor** | |
| **Description** | Customer provides booking details to request booking. |
| **Pre-Condition** | Customer must be logged in to the system. |
| **Success Scenario** | Booking request saved to the database and forwarded to the employee. |
| **Failure Scenario** | Booking request not saved to the database and not forwarded to the employee. |

*Table 3.4: Use Case Description for Booking Cancellation Prediction*

| | |
|---|---|
| **Use Case ID** | UC-04 |
| **Use Case Name** | Booking Cancellation Prediction |
| **Primary Actor** | Employee |
| **Secondary Actor** | |
| **Description** | System predicts whether customer is likely to cancel booking or not. |
| **Pre-Condition** | Customer requests booking and employee must be logged in to the system. |
| **Success Scenario** | Prediction result is saved to the database. |
| **Failure Scenario** | Prediction result not saved to the database. |

### ii. Non Functional Requirements

Non-functional requirements define how a system should perform its functions rather than what functions it should perform. Non-functional requirements are frequently linked to characteristics like performance, security, usability, and maintainability. They are essential in ensuring that the system adheres to specific standards and limits.

The non-functional requirements of the system are as follows:

- **Security:** The system allows only the registered users to log in to the system and any other users are not allowed to access the system functionality without logging in. Customers and employees in the system are provided with different privilege levels.
- **Usability:** The system has simple, responsive, and navigable interface that is easy to use.
- **Maintainability:** The system is built using object-oriented approach which can be easily modified according to changing requirements and the documentation also adds to the maintainability of the system.

## 3.1.2 Feasibility Analysis

Feasibility analysis typically considers various aspects, including technical, economic, legal, operational, and scheduling factors. Feasibility analysis is the process of evaluating the practicality and viability of a proposed project or system. It assesses whether the project is worth undertaking and whether it can be successfully completed.

### i. Technical Feasibility

Technical feasibility involves evaluating whether the proposed system can be developed and implemented with the available technology and resources. During this process, the availability and quality of data needed for training and testing the prediction model are assessed. For this project, the data includes historical booking data, and customer information which is quite comprehensive, with minimal inconsistencies. The team

members are also well-versed in the technologies used, so this project is technically feasible.

### ii. Operational Feasibility

Operational feasibility evaluates whether the proposed project can be effectively integrated into an organizations existing operations. Assessing the compatibility and interoperability of the newer system is an important step in operational feasibility. Since this system is developed for hotels to effectively manage the risks that come with booking cancellations, it is safe to say that it fits perfectly with the organizations strategic goals, and processes. The system can be easily integrated with current systems and databases. Hence, this system is operationally feasible.

### iii. Schedule Feasibility

Schedule feasibility is an important aspect of project planning and management. Schedule feasibility involves determining whether the proposed project can be executed within the predefined time schedule. By analysing the resources, timeline, and risks, this project was deemed to be feasible in schedule and can be successfully completed in given time.

The following figure shows the WBS for the project:

### 3.1.3 Analysis

In the analysis phase, requirements of the system are structured. Data modelling using ER Diagram is used to structure the requirements.

The following figure shows the ER diagram of the system.

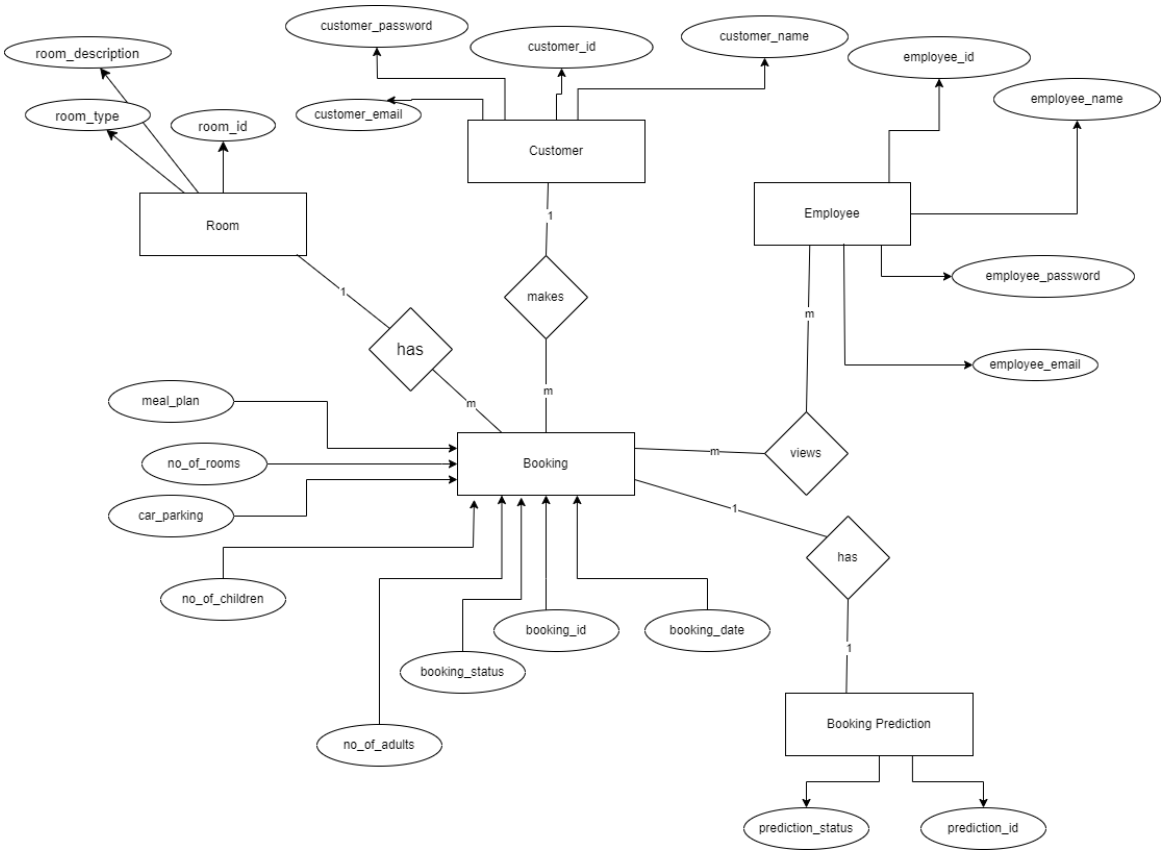


*Figure 3.2: ER Diagram*

# CHAPTER 4: SYSTEM DESIGN

## 4.1 Design

The design phase is a crucial stage where the overall structure and architecture of the system are planned and defined. The main focus of this phase is to transform the requirements into a blueprint which can then be used to build the actual software. Class diagram, activity diagram, and sequence diagrams are used to illustrate the design and workflow of the system.

### 4.1.1 Class Diagram

Class diagrams are a type of Unified Modelling Language (UML) diagram used to visually represent the structure of the system, and the relationships between its different classes. Class diagrams provide a high-level over view of the system, focusing on the classes, their attributes, methods, and the associations between them.

This system has five classes namely, Customer, Employee, Room, Booking, and Booking Prediction. The Customer and Employee classes have attributes such as id, name, email, and password. Room class has id, type, and description as its attributes. Booking has attributes like id, date, status, number of adults, number of children, car parking, number of rooms, and meal plans. Booking has a one-to-many relation with the customer. Finally, the Booking Prediction class has id, and prediction status.

The following figure shows the class diagram of the system.

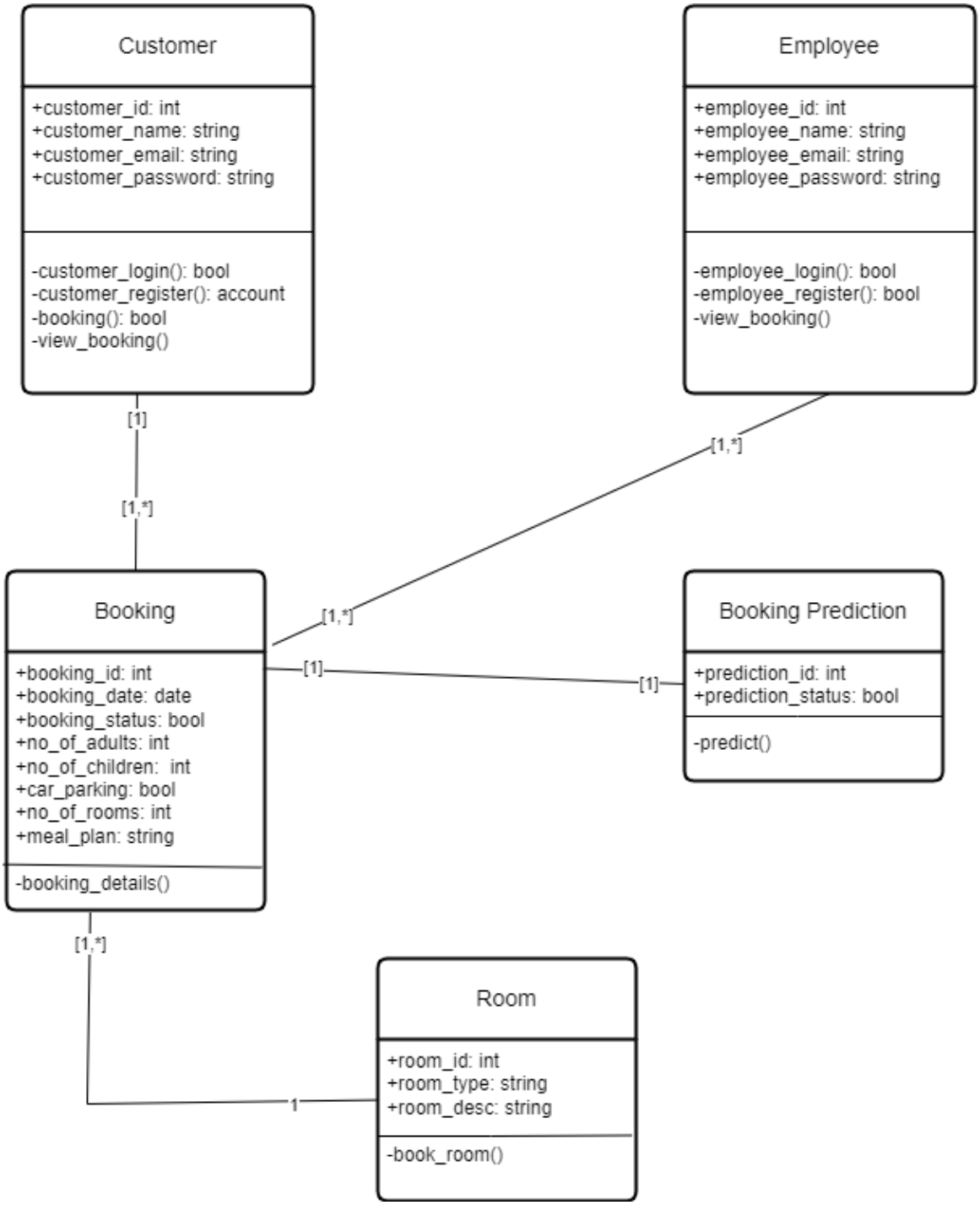


*Figure 4.1: Class Diagram*

### 4.1.2 Activity Diagram

Activity diagrams are also a type of UML diagrams, used to represent the flow of activities within a system. They are used for modelling business processes, workflows, and use cases. Activity diagrams illustrate the dynamic aspects of a system like the sequence of actions and the flow of control.

The following figure shows the activity diagram for booking a room.

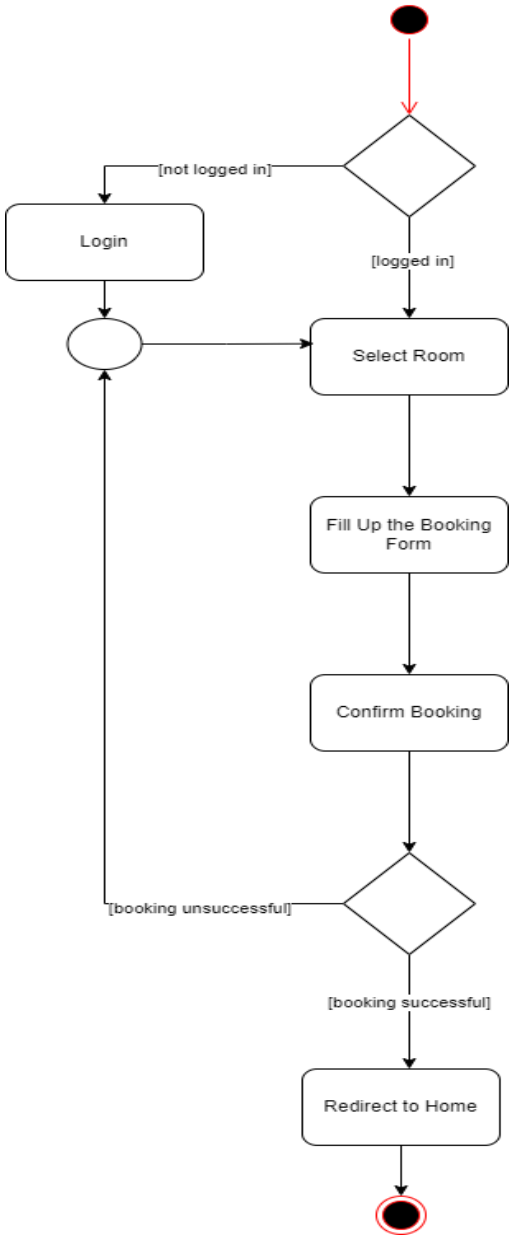


*Figure 4.2: Activity Diagram*

### 4.1.3 Sequence Diagram

Sequence diagrams, like Activity diagrams, also represent the dynamic aspects of the system. Sequence diagrams are used to illustrate the interactions between objects and components over time. They emphasize on the chronological order of interactions and messages exchanged between the different entities within the system.

The following figure shows the sequence diagram for predicting a booking cancellation.

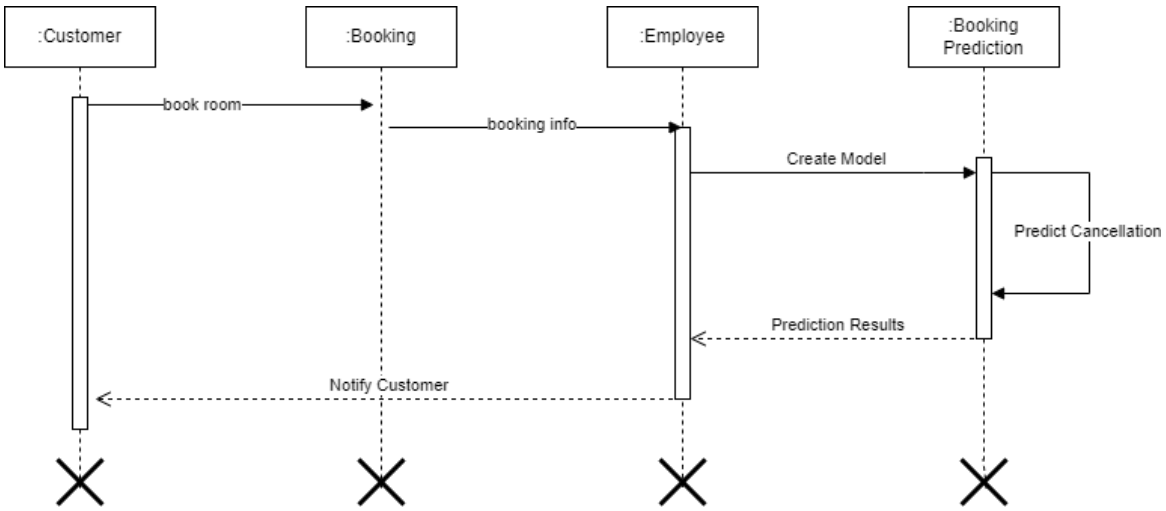


*Figure 4.3: Sequence Diagram*

## 4.2 Algorithm Details

# References


[1] A. J. S.-M. M. P. Eleazar C. Sanchez, “Identifying Critical Hotel Cancellations using Artificial Intelligence,” International Journal of Hospitality Management, 2020.

[2] Y. Lin, “Research on the Influencing Factors of Cancellation of Hotel,” Department of Automation and electrical engineering, Jinan University, Shandong, 2023.

[3] R. Mytnik, “Predicting Hotel Booking Cancellations,” *Analytics Vidhya,* 2021.

[4] S. T. M.R. Martinez-Torresa, “A machine learning approach for the identification of the deceptive reviews,” University of Seville (Spain), 2019.

[5] A. d. A. L. N. Nuno Antonio, “Big Data in Hotel Revenue Management: Exploring Cancellation Drivers to Gain Insights Into Booking Cancellation,” Cornell Hospitality Quarterly, 2019.