A PROJECT REPORT
On

HOTEL CHECK-IN ABANDONMENT FORECASTING MODEL

Submitted by

SOHEN MONDAL[RegNo:RA2011028030010]
AKSHAY DAWAR[Reg No:RA2011028030017]
AYUSH PAL[Reg No:RA2011028030021]

Under the guidance of
Ms. Surbhi Jain
(Assistant Professor, Department of Computer Science &

Engineering)*in partial fulfillment for the award of*

the degree of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE & ENGINEERING

of

FACULTY OF ENGINEERING AND TECHNOLOGY

SRM INSTITUTE OF SCIENCE & TECHNOLOGY, NCR CAMPUS

NOV 2023

SRM INSTITUTE OF SCIENCE & TECHNOLOGY

(Under Section 3 of UGC Act, 1956)

BONAFIDE CERTIFICATE

Certified that this project report titled "**Hotel Check-in Abandonment Forecasting Model**" is the bonafide work of "**AKSHAY DAWAR [RA2011028030017], SOHEN MONDAL[RA2011028030010], AYUSH PAL[RA2011028030021]**", who carried out the project work under my supervision. Certified further, that to the best of my knowledge the work reported herein does not form any other project re- port or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

SIGNATURE

SIGNATURE

Ms Surbhi Jain
GUIDE
Assistant Professor
Dept. of Computer Sciene & Engi-
neering

Dr. R.P.Mahapatra
HEAD OF THE DEPARTMENT
Dept. of Computer Science & Engi-
neering

Signature of the Internal Examiner

Signature of the External Examiner

ABSTRACT

- The "**Hotel Check-in Abandonment Forecasting Model "** helps to address the issue regarding the rate at which Hotel Rooms are cancelled. To address this issue, the above project uses Machine Learning to implement and predict when the rooms are cancelled.

- The project demonstrates how **predictive analysis** can contribute to synthesizing and predicting booking cancellations. This uses different types of algorithms such as **Regression, classification** etc. to classify data and evaluation matrices to separate categorical data into specific types.

- By doing a little formatting the required result can be obtained. This prevent poor room management by hotels and improves the client experience.

ACKNOWLEDGEMENTS

- I would definately like to express my deepest gratitude to my guide, Ms Surbhi Jain for her valuable guidance, encouragement, caring and timely help by providing me with an excellent atmosphere for doing research.

- All through the work, in spite of her busy schedule, she has extended cordial support to me for completing this project.

Sign.

TABLE OF CONTENTS

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

- Cancellations play a significant impact on demand and supply chain in the hotel industry.

- To reduce the effect of booking cancellations hotel management, provide with many tough terms which make the task more difficult which in return increases the work load of the staff.

- Therefore, it is important for the hotels to reduce the amount of cancellation so as to maximize the amount of profit made by the organization.

- The best way to do so would be using machine learning to make predictions beforehand and avoid any unnecessary problems afterwards.

- People can benefit from the project by making   predictions for their personal   use   cases.

CHAPTER 2

LITERATURE SURVEY

2.1 Tools used:

2.1.1 Python Language:

- Python is a general-purpose, high-level and it is quite popular these days. Extensive Data analysis and machine learning works can be done through it easily as compared to other languages such as c/c++,java,javscript etc.

- Python language is currently used by most techinacl organizations.

2.1.2 Jupyter Notebook:

- Jupyter Notebook is a latest interactive web based ide for python programming and testing.

- It improves workflow and efficiency.

2.1.3 Libraries Used:

Matplotlib version 3.0.3:

- Matplotlib is a 2D plotting based library which is often used in data science projects.

- It can easily generate histograms, plots, bar charts, error charts and scatterplots etc. It also also low code in nature as compared with other libraries.

Seaborn version 0.9.0:

- Seaborn is a library thatis built in conjunction to the matplotlib library.

- It aims to make graphs and plots a central part of the data science project.

Sklearn version 0.20.3:

- Scikit-learn (Sklearn) is a comprehensive library used for machine learning processes.

- In itself it is a compilation of several machine learning modules which can be easily imported and applied to a project.

NumPy version 1.24.3:

- NumPy is also known as Numerical Python and it was created in the year 2005.

- It delivers operation on linear algebra, arrays, transform series and matrices. It is fast and efficient(runtime) as it is built on c++ platform.

Pandas version 2.0.1:

- Pandas are used for working with datasets and data frames.

- It was created in the year 2008 and since then reading cleaning data using pandas has become much popular.

2.2 Data Science Concepts Used:

Machine Learning (ML):

- Machine learning is big term which comprises solving problem using various algorithms.

- A machine is given exposure to the data and problem statement and then we try to apply machine learning on it.

Hypothesis Testing:

- It is a way that is used to make stat based decision by using data and results.

Data Visualization:
- Data visualization is the representation of data via use of various graphs, plots and charts. These visuals are quite handy for understanding the underlying aspects.

Data Cleaning:

- It is one of the most important phase of machine learning. Fixing and filling values which are misleading are a part of it.

Feature Engineering:

- It is the big process of using any sort of domain language to crete a predictive model base.

- The major goal of it is to improve the performance of the ML model.

2.3 Problem Statement

The main objective of this project is to reduce the amount of cancellation loss suffered by the hotel a various hospitality services and to maximize the profit of the organization. It has come to the notice of the hotel agency that from year on base there has been increased trend on the cancellation of Room booked which leads to loss and is now becoming a norm that the industry must take these precautions before finalizing their operational cost.

The result of cancellation of bookings in the hospitality industry causes:
● Labor allocation and Financing
● Revenue Loss
● Reduces the reputation of the hotel organization
● decisions Overbooking of Rooms

2.4 Project Planning

Requirements:

● Which Market segment is affected by Cancellation?

● How lead time affect the cancellation of Rooms?

● What is rate of Booking and cancellation in various countries?

● What is the reason for cancellation?

● Which machine learning algorithm can provide the best accuracy for the above model?

Steps To Be Followed:

● **Data Cleaning**:

- Data Cleaning is the process of fixing the raw data before we use it.

- It is one of the most important phase of machine learning. Fixing and filling values which are misleading are a part of it.

● **Exploratory Data Analysis**:

  - Exploratory data analysis (EDA) is a method used to explore the data nad it's underlying features.

● **Feature Selection:**

  - It is the big process of using any sort of domain language to crete a predictive model base.

  - The major goal of it is to improve the performance of the ML model.

● **Model Building**:

  - A machine is given exposure to the data and then we try to apply machine learning on it by creating a model based on it.

Usually There are 3 types of machine learning models:
  1. Reinforcement Learning.
  2. Supervised Learning.
  3. Unsupervised Learning.

CHAPTER 3

SYSTEM ANALYSIS

3.1 Design Constraint:

Jupyter Notebook:

The Jupyter Notebook is the original web application for creating and sharing computational documents. It offers simple, streamlined, document-centric experience. Project Jupyter is a project to develop open-source software, open standards, and services for interactive computing across multiple programming languages.

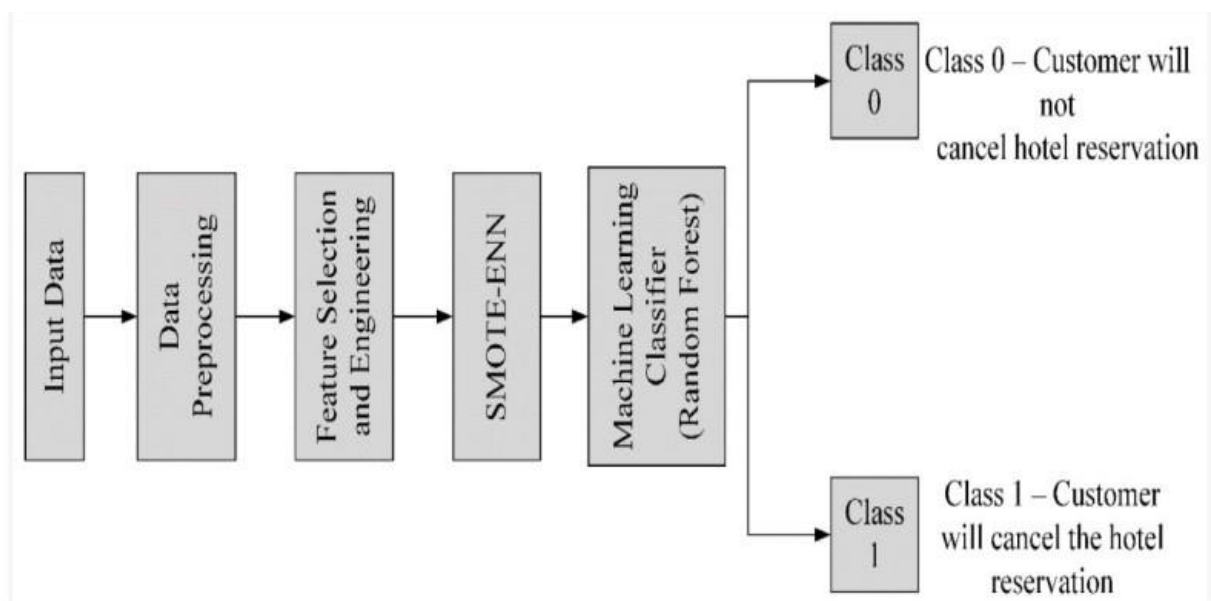## 3.2 System Architecture OR Block Diagram

Fig 1: Block Diagram Representation Of Conceptual Methodology

CHAPTER 4

SYSTEM DESIGN

4.Implementation

During the project development, the following steps were taken:

4.1 Methodology

Data Loading and Reading: In this step, we loaded the dataset into the environment and read the data using Python's Pandas library.

Data Cleaning and Preprocessing: We removed missing values, duplicates, and outliers from the dataset. We also performed data imputation for missing values using different techniques like mean imputation, mode imputation, and regression imputation. Additionally, we scaled the numerical variables using standard scaling to ensure that all variables have a similar scale. Furthermore, we converted categorical variables into numerical variables using one-hot encoding.

Exploratory Data Analysis: In this step, we analyzed the origin of guests, price paid per night by guests, busiest months for bookings, month with the highest Average Daily Rate (ADR), and whether bookings were made for weekdays, weekends, or both.

Feature Engineering: We created new features like total number of guests, total number of nights booked, and average price per night.

Feature Encoding: We transformed categorical variables into numerical variables using one-hot encoding.

Graphs and Plots: Additionally we analyzed the origin of guests, price paid per night by guests, busiest months for bookings, month with the highest Average Daily Rate (ADR), and whether bookings were made for weekdays, weekends, or both i.e the columns of the dataset with the help of matplotlib and seaborn.

Feature Selection: We selected important features using correlation and univariate analysis.

Model Building: We built a machine learning model using Logistic Regression to predict weather the booking will get cancelled or not.

Model Cross-Validation: We performed cross-validation to evaluate the performance of the machine learning model.

CHAPTER 5

CODING, TESTING

5.1 Testing OR Verification Plan

We tested the machine learning model using various test cases to ensure that the model is performing well on different datasets. We also performed hyperparameter tuning to improve the performance of the model.

Test	Test Case Title	Test Condition	System Behavior	Expected Result
T01	Data Loading Test	The data loading function is called with a valid input file path.	The function should read the input file and load the data into memory.	The data is loaded without errors and is available for further processing.
T02	Outlier Detection Test	The outlier detection function is called with a sample dataset containing known outliers.	The function should identify the outliers and mark them for removal.	The identified outliers should be removed from the dataset and the remaining data should be suitable for further processing.

5.2 Result Analysis OR Screenshots

In this subsection, we present the output of the experiment or project in terms of some graphs, plots. We have generated various plots like scatter plot, box plot, heatmaps, and correlation matrix to visualize the data and understand the relationships between variables.
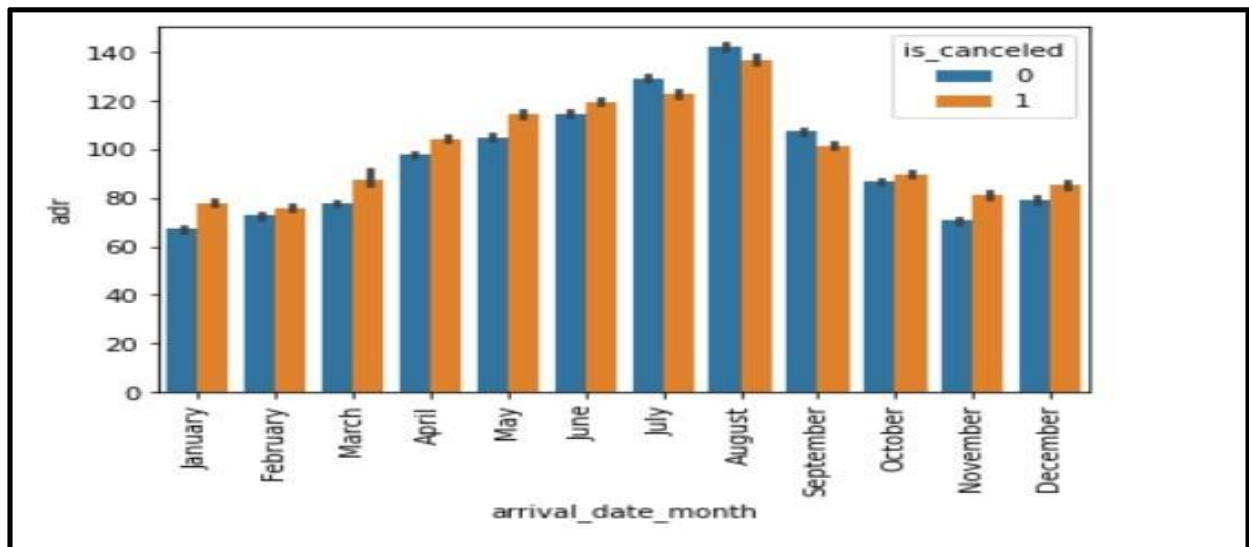

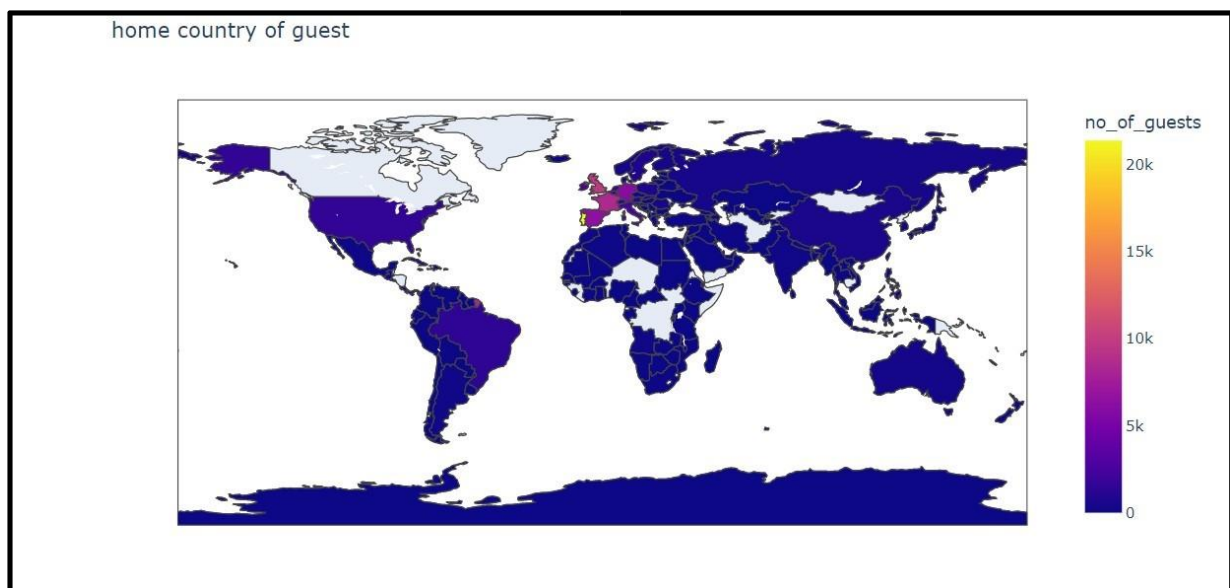**Fig 1 : Graphic Visualization of Average daily Rate and Arrival date month**

Fig 2: Graphics Visualization Of Guests

```
# Train the model with logistic regression
from sklearn.linear_model import LogisticRegression
classifier = LogisticRegression()
classifier.fit(x_train, y_train)
y_pred=classifier.predict(x_test)

    /usr/local/lib/python3.10/dist-packages/sklearn/linear_model/_logistic.py:458: ConvergenceWarning: lbfgs failed to
    STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

    Increase the number of iterations (max_iter) or scale the data as shown in:
        https://scikit-learn.org/stable/modules/preprocessing.html
    Please also refer to the documentation for alternative solver options:
        https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
      n_iter_i = _check_optimize_result(


# check for accuracy score of regression algorithm
from sklearn.metrics import accuracy_score
print(accuracy_score(y_test,y_pred))

    0.7952089789764637

This function computes subset accuracy: the set of labels predicted for a sample must correspond to y_test.
```

Final Output Of ML Model (Logistic Regression)

5.3 Quality Assurance

We followed the guidelines and standards set by our university to ensure the quality and level of our project. We also obtained feedback from our team members and our project guide to improve the project's quality. Quality was improved gradually with passage of time.

5.4 Standards Adopted

Design Standards

- Defined project objectives and scope clearly
- Used appropriate design patterns and principles
- Followed industry-standard frameworks and architectures
- Used modular and scalable design approaches

Coding Standards

- Used meaningful variable and function names
- Commented the code to improve its readability and maintainability
- Used consistent coding style throughout the project
- Used version control systems to manage code changes
- Used automated code analysis tools to detect errors and maintain quality

Testing Standards

- ISO/IEC 9126: Software engineering - Product quality
- ISO/IEC 12207: Software life cycleand processes
- IEEE 1061: Software Quality Metrics Methodology
- Defining test cases and test scenarios
- Conducting unit testing, integration testing, and system testing
- Performing regression testing and acceptance testing
- Tracking defects and issues using a defect tracking tool
- Generating test reports and documentation

CHAPTER 6

CONCLUSION

- Logistic Regression Has the Best Accuracy Among All Algorithms That We Tried from all the evaluation matrix to predict hotel cancellation classification case, we see that Logistic Regression has the best accuracy when it comes to predicting hotel cancellation based on certain features (80%).

- Logistic Regression works in such a way that it, uses a sigmoid function to classify instances as true and false. Hence it is quite effective for binary classification problems.

$$f(x) = \frac{1}{1+e^{-\beta x}}$$

Sigmoid
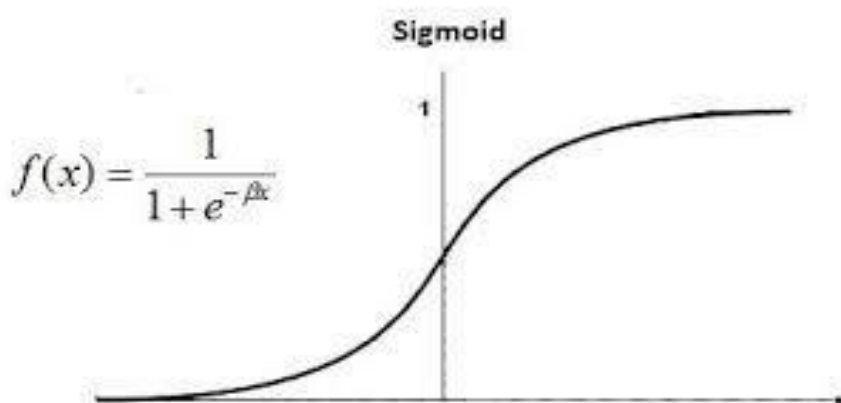
- This model enables hotel managers to reduce the losses and focus on the underrated aspects to improve their sales.

CHAPTER 7

FUTURE ENHANCEMENT

7.1 Future Scope

A hotel check-in abandonment forecasting model can provide valuable insights for hotel management to optimize their operations and improve guest experience. Here are some potential future scopes of such a model:

1)**Real-time optimization:** The model can be integrated with real time data feeds to allow hotel staff to make informed decisions on the spot. For example, if the model predicts high check-in abandonment rates during a particular time of day, the staff can allocate more resources to that time to ensure a smooth check-in process.

2)**Personalization:** The model can be customized to take into account the guest's history, preferences, and behavior to provide a more personalized experience. For instance, if the model predicts that a guest is likely to abandon check-in due to a long queue, the hotel can offer them a personalized check-in service.

3)**Revenue optimization:** The model can help hotels optimize their revenue by predicting the optimal number of rooms to overbook. By considering the historical data and trends, the model can suggest the right amount of overbooking without risking the guest experience.

4)**Customer loyalty:** By reducing check-in abandonment rates, the model can help hotels improve customer loyalty. When guests have a seamless and efficient check-in experience, they are more likely to return and recommend the hotel to others.

5)**Continuous improvement:** The model can help hotels continuously improve their operations by providing insights into the causes of check-in abandonment

Overall, a hotel check-in abandonment forecasting model has a lot of potential for enhancing guest experience and optimizing hotel operations. As technology advances and more data becomes available, the scope of such a model will only increase.

REFERENCES

1.Title: Hotel Check-In Abandonment Forecasting: A Machine Learning Approach
Author: Mohammed Al-Momani
Publication: The International Journal of Engineering and Science, 2021

2. Title: Predicting Hotel Check-In Abandonment Using Machine Learning
Author: Alireza Jalali
Publication: Journal of Tourism and Hospitality Management, 2022

3. Title: Hotel Check-In Abandonment Forecasting Using a Long Short-Term Memory Neural
Network
Author: Alireza Jalali and Mohammadreza Hosseini
Publication: arXiv preprint arXiv:2209.10428, 2022