Hotel Booking Cancellation/Demand Prediction

Sneha Jayaraman
5th Semester
Department Of Computer Science and Engineering
PES University
Bengaluru, India
Sneha.jayraman@gmail.com

Hemanth Alva
5th Semester
Department Of Computer Science and Engineering
PES University
Bengaluru, India
Hemathalva2708@gmail.comthrupthijiyo22@gmail.com

Thrupthi H M
5th Semester
Department Of Computer Science and Engineering
PES University
Bengaluru, India
thrupthijiyo22@gmail.com

I. INTRODUCTION

Hotel booking cancellations hurt the hospitality industry the most in terms of revenue and reputation. Most hotels implement a rigid cancellation policy to do away with the negative impact on the top line figures. This strategy, however, sacrifices the flexibility available to the customers and results in customer dissatisfaction. Anytime cancellation policy, on the other hand, will result in the hotel losing out on the profits of rooms that could have otherwise been sold if the cancellation was made well in advance. More recently, online travel agencies have become a common mode of booking that allows customers to book and cancel at any time, free of cost. This encourages overbooking. The hotels now face uncertainty and must devise strategies to profitably sell the rooms ensuring that they do not lose out on potential customers. The hotels are therefore tasked with creating a balance between profits generated and customer satisfaction.

One way to achieve this would be to devise revenue management strategies such as limiting the number of days for the cancellation to be a predefined fixed day from the arrival day i.e. for example 5 days before the arrival day. This, however, would not necessarily comply with the competitor strategy and the customers may prefer hotels that are more lenient in their policy. Demand in hotel rooms is dependent on various factors such as seasonality, holiday occurrences in addition to the facilities provided by the hotel itself. Predicting periods of high demand will help the organisation be better equipped. Further, implementation of dynamic pricing i.e. increasing the average daily rate on the high demand days would be possible only if the hotel could accurately forecast demand. The solution, therefore, would be to leverage the capabilities of data analytics and machine learning to identify customer behaviour and improvise demand-management decisions.

The bigger picture involves understanding customer groups and categorising customers that are likely to cancel their booking on a given day. Besides, identifying other features such as the distance of travel, booking mode, holidays that could affect the cancellation rate plays an important role in predicting demand. Identifying predictors that most impact the demand will help hotels and revenue managers make better-informed decisions. Providing perks or incentives to customers is a way of encouraging to book more. Using data science can therefore provide a huge competitive business advantage to the hospitality industry.

II. LITERATURE REVIEW

Predicting or forecasting booking cancellations is mostly considered a problem of regression since the parameter used to predict is the rate of cancellation rather than the individual booking cancellation probability themselves. [1] however considers this to be a case of classification and uses binary classification models such as boosted decision trees, decision forests, decision jungle, locally deep support vector machine and neural networks to forecast whether a given booking will be cancelled or not. A prediction model was devised for each hotel as the actions that cause a booking cancellation could be attributed to the services specific to a hotel in terms of its location or even the discounts that a hotel offers. The number of false positives (falsely equated to a cancellation) was chosen as an important parameter since the hotel would not like to lose potential customers not cancelling their bookings. The Decision Forest model performed the best overall with well over 90% accuracy. The models, however, predict which booking is likely to be cancelled on a daily basis. Moreover, the features that capture seasonality and trend were not used for modelling.

Autoregressive Integrated Moving Average Model, EMD, EEMD, EEMD-ARIMA models were used to predict the hotel occupancy or demand thereby taking into consideration the time series trend that influenced the bookings [2]. Time series models provide for a practical implementation taking into account the seasonal components. ARIMA model produced considerable fit and a better forecasting accuracy than the EEMD-ARIMA for the study. A modified EEMD-ARIMA performed better with medium-term forecasting. For a hotel, the exact number of booking cancellations per day is of lesser interest. The degree of overall increase or decrease in cancellations weekly would be more meaningful from a business standpoint. Identifying the trend over a fixed period might

help the hotels make better-informed decisions in their revenue strategy and cancellation policies [2].
[3] models the probability of cancellation as a probit model that takes into account several features that have an association with the target variable. This theoretical model shows the magnitude of the relationship between seasonal components and other external features on the booking cancellation. The empirical model can account for the large heterogeneity in the behaviour of customers and thereby determine the influencing attributes that could result in a booking cancellation that would otherwise be thought of as inevitable. A total of 6 hypotheses were formulated. These include possible reasoning to booking cancellations such as the impact of country of origin, mode of booking (online or onsite), the distance of travel, seasonality, size of booking, etc. Descriptive statistics is important in understanding the problem but only provides a theoretical framework on the most important exploratory variables.

III. PROPOSED PROBLEM

Cancellation prediction can be thought of as a probabilistic prediction problem wherein each booking can be classified as cancelled or not cancelled based on classification algorithms such as Logistic Regression, Decision Trees, Random Forest, XGBoost algorithm. Feature selection, dimensionality reduction techniques could be applied to identify the most important predictors affecting the cancellation probability. Formulating a theoretical model that identifies relationships between exploratory variables is important for revenue managers in terms of delivering on a profit margin. The rate of cancellation could be accurately forecasted using the supervised learning techniques above.

To account for the seasonality in the data and the impact of time in terms of seasons of fixed periodicity, an ARIMA model could be developed. Further, a hotel may only be interested in the approximate trend of cancellations for a particular period rather than the exact estimate of cancellation rate calculated for a day, let alone the classification of each booking in terms of the whether the booking is likely to be cancelled or not. Thus, a weekly approximate for the cancellation rate could be calculated that would be more beneficial for devising a revenue-management decision. Likewise, demand could be estimated on a weekly basis to check for when a hotel is probable to get a disproportionately large number of bookings.

The dataset chosen describes hotel demand/bookings data. This dataset comprises two hotels (a resort hotel, a city hotel).It has 32 columns wherein each observation indicates a feature related to hotel booking such as arrival date, lead time(number of days between booking and check in), number of adults, children, babies, type of meal booked, country of origin, previous cancellations, previous bookings not cancelled by the user, type of room reserved, booking changes made after the booking, deposit type, travel agent that made the booking, ADR(Average Daily Rate), number of special requests, number of days the booking was in the waiting list, customer type and so on. The motive here is to find the relationship between all of these features and booking cancellation, and help increase the ADR (revenue) of the hotel. We look to identify the answers for the following:

- How many bookings were cancelled in our dataset and does this depend on the time of the year of booking?
- In a specific year, which month is the busiest with the least number of cancelations? Could we be able to predict when the hotel was likely to receive a disproportionately high number of special requests?
- What would be the optimal length of stay which had the least cancellation rate / What is the optimal length of the stay in order to get the best daily rate?
- Is the number of booking changes/amendments made to a booking deterministic of a cancellation?
- How can we get a better understanding of the relationship between a customer and the number of customizations asked for?
- How does price vary per night over the year?
- Does a customer with a booking from a travelling agent influence the cancellation and which agency/company results in a large number of cancellations?
- Could we cluster/segment bookings based on their features?
- Do repeated guests show a lower chance of booking cancellation and does a new customer tend to be more vigilant to cancellation?

IV.HOW IS IT DIFFERENT FROM EXISTING SOLUTIONS

Hotel occupancy forms a major part of the hospitality industry. Cancellation prediction has been seen as an inevitable and unexplainable event. However, recent research has focussed a lot on the booking cancellations and booking demand as a function of features that reflect customer behaviour that could lead to a cancelled booking. Past history can indeed be used effectively to model demand. However, the majority of research ignores the time-varying component of cancellation that could be due to demand fluctuations as in the case of holidays or seasonal events affecting the cancellation rate.

This study, therefore, aims to develop models that take into account the time-varying trend in addition to predictors. Moreover, the cancellation rate on a weekly basis rather than a day-to-day basis that helps identify the overall upward or downward trend will help the hotels better gauge

future hotel room demand thereby providing them with the opportunity to resell the room. The exact rate, in this case, will be of lesser importance than the overall trend.

REFERENCES

[1] Antonio, Nuno & De Almeida, Ana & Nunes, Luís. (2017). Predicting Hotel Booking Cancellation to Decrease Uncertainty and Increase Revenue. Tourism and Management Studies. 13. 25-39. 10.18089/tms.2017.13203.

[2] Muzi Zhang Junyi Li Bing Pan and Gaojun Zhang. (2018). Weekly Hotel Occupancy Forecasting of a Tourism Destination. MDPI

[3] Falk, Martin & Vieru, Markku. (2018). Modelling the cancellation behaviour of hotel guests. International Journal of Contemporary Hospitality Management. 10.1108/IJCHM-08-2017-0509.