



Survey on Hotel Booking Cancellation Prediction Using Machine Learning Model

Prajwal Shiwal¹, Dhruv Morey², Harshad Shivankar³, Suraj Jagtap⁴,
Prof. Sushadevi Adagale⁵

^{1,2,3,4,5}Computer Engineering, Trinity Academy of Engineering, Pune, India

Abstract

The cancellation of bookings puts a considerable strain on management decisions in the case of the hospitality industry. Hotel Booking Cancellation predictions are a vital tool for revenue management performance. However, in recent times, thanks to the availability of considerable amount of computing power and through machine learning (ML) approaches, it has become possible to create more accurate models to predict the cancellation of bookings compared to more traditional methods. Previous studies have used several ML approaches, such as Support Vector Machine (SVM), and Decision Tree (DT) models for predicting hotel cancellations. So In Nutshell to fill this gap and investigate how these problem barriers can be solved one must need to understand the different factors based on hotel management system and the customers requirement and also check for whether those factors are implemented or not , An automated Machine Learning (ML) model can be designed using hybrid approach by understanding, analyzing and executing different Machine Learning Algorithms such as Naive Bayes (NB),K-Nearest Neighbors Algorithm (K-NN),Random Forest Algorithm and Logistic Regression (LN),etc.

I. INTRODUCTION

Tourism is one of the most expanded diligence in the world and its significance in the global frugality. The trip assiduity is aiding work with bettering lives for a huge number of people and changing entire networks. Reserving cancellation is a crucial aspect of hotel profit operation as it affects the room reservation system and guest allocation system. In the hospitality assiduity the description of profit administration is “ making the right room available for the right person and the genuine price at the apparent time via the right rotation medium”. Considering that



lodgings (Hotels) have an established number of rooms, and that they offer them as a perishable item to give the right room to a suitable individual, lodgings have to admit movables ahead of time. Booking is a kind of an agreement between a lodging and its guests, and it gives guests the right to cancel an agreement. For Hotel, bookings in advance are the main index of a hostel's cast performance. Still, cancellations impact Hotels further than guests, as a hostel should have rooms for guests who admire their bookings but, at the same time, it struggles financially when a customer cancels a booking or does not show up. A booking cancellation occurs when a customer closes their contract before their entry, while a no-show is when a customer doesn't inform the lodging of a change in plans and fails to check in. Still, booking may also be canceled due to some other reason similar as bad rainfall, holiday cataloging, unforeseen illness, change in meeting place, and numerous others. As a result, cancellations have a adverse effect on demand administration choices within a profit administration frame. Predicting a cancellation before hand can profit both, the hostel and the guests to a great extent. As reserving cancellations are frequently resolved by regression and classification problems, it's important to understand when to settle on between these two styles, as an illustration, when the only end is to estimate cancellation rates, also it's considered a regression problem still, when the end is to estimate the liability of a booking being canceled it should be considered as classification system. In machine literacy (ML), supervised literacy is naturally partitioned into two feathers of problem "regression", when an affair is quantitative (e.g., stock request vaticination), or " classification", when an affair is categorical or separate (e.g., soothsaying in the case of hostel bookings that show whether a client " will cancel booking" or " will not cancel booking").

II. LITURATURE SURVEY/BACKGROUND

Booking cancellation could be a well-known issue in revenue administration, and it's applicable to the industry and, most importantly, to the hospitality industry. Customer's increasing interest within the net has changed the way within which they buy or search for any service. Current customer behavior features a substantial influence on contemporary research on the matter of booking cancellations, particularly that associated with the results of cancellations on revenue and inventory allocation, similarly as on cancellation and overbooking policies. The primary paragraph under each heading or subheading should be flush left, and subsequent paragraphs should have a five-space indentation. The importance of forecasting in Revenue Management & Demand Forecasting is identified united of the focuses of research in Revenue Management, within the hospitality industry (rooms division), revenue management general definition was adapted to "making the proper room available for the correct guest and also the right price at the proper time via the proper distribution channel", within the hospitality industry accepts bookings earlier. These bookings represent a contract between the customer and also the hotel. This contract gives the customer the correct to use the service within the future at a settled price, but most of the days, with an choice to cancel the contract before the service provision. Although advanced bookings are considered the leading predictor of a hotel's forecast performance, this feature to cancel the service puts the chance on the side of the hotel, resulting in the hotel having to confirm that it's rooms to customers who honor their bookings. However, a cancellation or a no-



show forces the hotel to endure the value of getting vacant rooms (although there are differences between no-shows and cancellations, for the aim of this study, both are treated as cancellations). a substantial number of studies are published on the topic of bookings cancellation and demand forecast. Nevertheless, a considerable part of these studies focuses within the airline industry, which although having some similarities to the hospitality industry, it's different. A second consideration is that almost all of those studies employ "traditional statistics" methodologies. Few of them profit of machine learning methodologies and techniques. The identical is additionally valid for research regarding the component of demand forecast to predict cancellations but only the studies of focus to the hospitality industry, and only use hotel specific data. This study, by using hotel specific data to develop a machine learning based model and deploy this technique during a prototype in an exceedingly real production environment, aims to increase a previous presented approach and present a case study showing that booking cancellation prediction is feasible, not only from theoretical standpoint, but also from an empirical one additionally.

III. PROPOSED WORK/SYSTEM

Very first we are going to collect some data and then perform the pre- processing and data cleaning on the data. In real world whatever data we get we always have it in the form of raw data and then we need to make our data ready for Machine Learning purpose. While considering the data from the Hotels, Rating or the number of Stars to the hotel can play a vital role in feature selection. People will choose the Hotels with maximum number of stars according to the ratings. So by using our prediction Model of Hotel Booking Cancellation the Hotels can take the certain precautions or the Safety Measures towards the factors which are causing the cancellation of the booking and improve their services towards customers which will create a good image of the hotel in the market, Eventually increasing the positive reviews and ratings. Then we are going to perform Exploratory Data Analysis to understand the data. Then we need to apply various techniques of Feature Engineering on our data. Such as:-Feature Encoding, Outlier Detection, Feature Selection. Feature selection is performed to select the imported features that have more influence on prediction, while feature engineering is performed to create other features from existing features, which can have a positive impact on classifier performance. After feature selection and engineering.

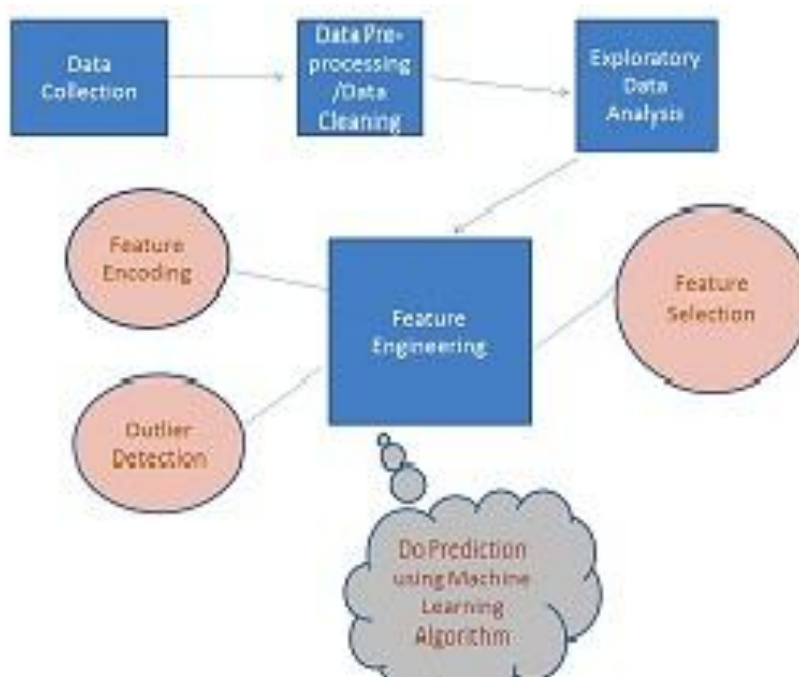


Figure presents a list of all the variables extracted.

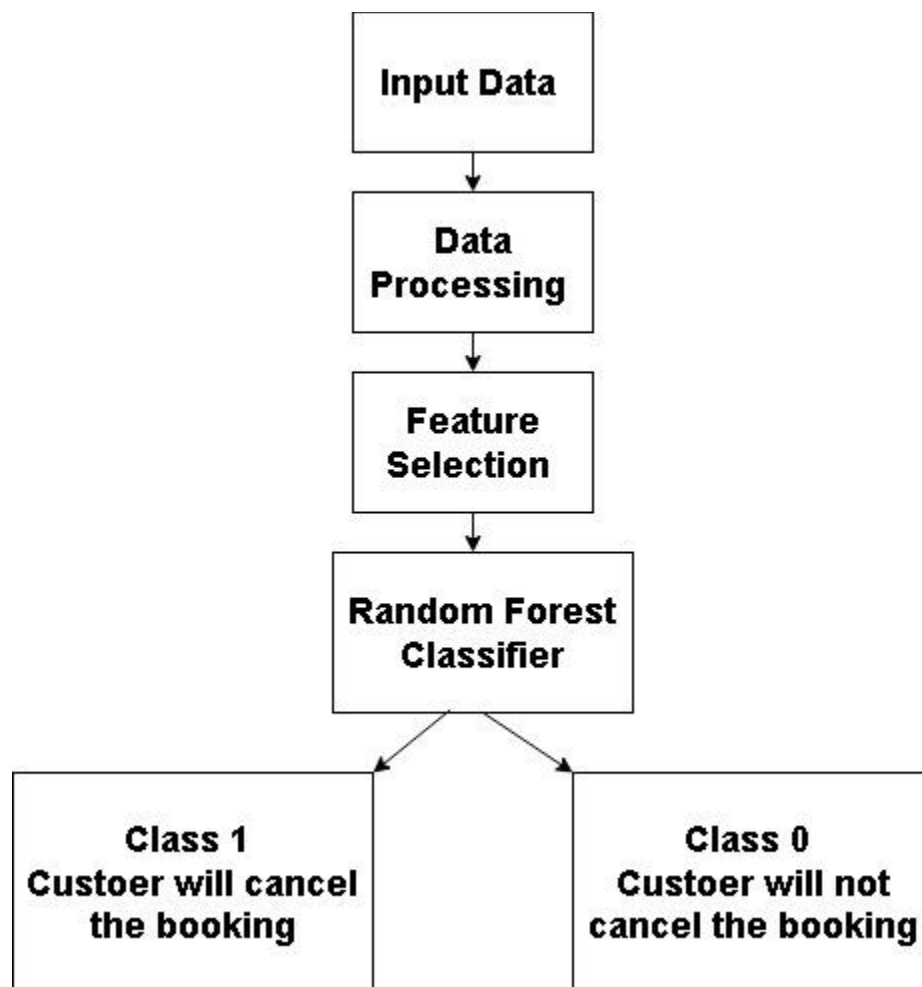
Name	Type	Description
ADR	Numeric	Average daily rate
Adults	Number	Number of adults
AgeAtBookingDate	Number	Age in years of the booking holder at the time of booking
Agent	Categorical	ID of agent (if booked through an agent)
ArrivalDateDayOfMonth	Numeric	Day of month of arrival date (1 to 31)
ArrivalDateDayOfWeek	Categorical	Day of week of arrival date (Monday to Sunday)
ArrivalDateMonth	Categorical	Month of arrival date
ArrivalDateWeekNumber	Numeric	Number of week in the year (1 to 52)
AssignedRoomType	Categorical	Room type assigned to booking
Babies	Numeric	Number of babies

Machine learning model

• Dataset Construction and Splitting Data for hotel forecasting has two aspects, one connected to when the booking was made and the other relating to the length of stay. This implies that a booking can be cancelled at any time between the time of booking and the planned arrival date (regardless of whatever cancellation rules hotels may have). As a result, a PMS database has three types of reservations at any one time:



- **Effective:** Reservations made at any point in time, with an arrival date that is less than or equal to the present date, and that have already checked-out or are currently checking-in.
- **Canceled:** Bookings made at any point in time, with an arrival date in any point in time (past or future), that were cancelled or did not show up.;
- **Future arrivals:** Bookings made at any point in time with any arrival date equal or greater than the current date, which cannot be cancelled until the projected arrival date, but can be cancelled until the present date. This signifies that these reservations are in a "unstable" status, meaning that some may be cancelled.





A. •Training Of the Model:-

- a) : The system trains daily with a data set of all reservations on the books, enabling it to find out with changes in bookings and changes of patterns that occur over time.
- b) : Each day, the system builds a replacement model and automatically executes hyper-tuning of parameters.
- c) : Performance is compared with the performance results of the previous seven days to get a choice that permits the model parameters to get replaced with new parameters or continued use of the previous parameters.
- d) : The predictions and performance results of the preceding days are stored in an exceedingly database for evaluation, and where applicable, reused as model elaboration features.
- e) : 50 percentage of the new bookings should be marked to perform the A/B Testing.

B. •Usability Of the Model:

- a) : A web-based platform with a visualization component should be accessible by the Hotel Reception anywhere and anytime.
- b) : Hotels should have a user login per staff to access the application. Every action taken by hotel staff should be logged.
- c) : Details of bookings that were identified as likely to get canceled for any particular date or upcoming days should be available for consultation.
- d) : Booking attributes which will be used for the identification of consumers should be displayed or recorded by the system (to enable research purpose usage).
- e) : The system should report the actions made toward bookings that were identified as likely to cancel to forestall their cancellation.
- f) : The system must provide the visualization of the model performance results daily.



g) : The system must provide the analysis of model predictions and effective performance results without disclosing the results of the A/B testing.

IV. CONCLUSIONS

The application of information science skills like data visualization, data processing, and machine learning, was possible to answer objectives of the research like Identify which features contribute to predict a booking cancellation probability. Application of information visualization and different data analysis techniques, together with the application of the different AI based Machine Learning Algorithms were understood. The development and deployment of the models in an exceedingly prototype tested in real-world conditions enabled the assessment of the system's relevance and predictability, other relevant roles of predictive analysis research. The results validated the worth of the system architecture design for running an automatic machine learning system that daily incorporates new data and utilizes previous prediction errors and hits for continuous improvement. Other of the important contributions of this study is that the development of a weighting system with training the model to boost the importance of newer observations and simultaneously learn from previous predictions. Additionally, this study also demonstrates how data-splitting method selection and domain knowledge in feature engineering are of paramount importance in machine learning modeling and therefore the influence for the advance of prediction models. Cancelled bookings are a major nuisance for hotels, and they are one of the leading causes of revenue and profit losses. As we've all seen, internet booking platforms are pushing more and more clients to book many hotels before deciding where they'll stay, contributing to the rise in cancellations we've all seen in recent years. We're witnessing changes in consumer behaviour as a result of the COVID-19 outbreak, but it's too early to make any predictions. We've already seen hotels reduce the number of days a client may cancel to encourage them to book more. So, while managing and forecasting cancellations was vital before the epidemic, it will be even more critical in the days ahead for the industry. With the advancement of research and data mining techniques, machine learning algorithms have improved.

V. REFERECNES

[1] Adil, M.; Ansari, M.F.; Alahmadi, A.; Wu, J.-Z.; Chakraborty, R.K. Solving the Problem of Class Imbalance in the Prediction of Hotel Cancellations: A Hybridized Machine Learning Approach. *Processes* 2021, 9, 1713. DOI:10.3390/pr9101713

- [2] Antonio, N, et al. 2019. An Automated Machine Learning Based Decision Support System to Predict Hotel Booking Cancellations. *Data Science Journal*, 18: 32, pp.1–20. DOI:10.5334/dsj-2019-032
- [3] Caicedo-Torres, W and Payares, F. 2016. A machine learning model for occupancy rates and demand forecasting in the hospitality industry. In: *Advances in Artificial Intelligence – IBERAMIA 2016*. Presented at the Ibero-American Conference on Artificial Intelligence. Cham: Springer. pp. 201–211. DOI:10.1007/978-3-319-47955-217
- [4] Antonio, N, de Almeida, A and Nunes, L. 2017a. Predicting hotel booking cancellation to decrease uncertainty and increase revenue. *Tourism Management Studies*, 13(2): 25–39. DOI:10.18089/tms2017.13203
- [5] Antonio, N, de Almeida, A and Nunes, L. 2017b. Using data science to predict hotel booking cancellations. In: Vasant, P and M, K (eds.), *Handbook of Research on Holistic Optimization Techniques in the Hospitality, Tourism, and Travel Industry*. Hershey, PA, USA: Business Science Reference. pp. 141–167. DOI:10.4018/978-1-5225-1054-3.ch006
- [6] Abe, N, Zadrozny, B and Langford, J. 2004. An iterative method for multi-class cost-sensitive learning. In: *Proceedings of the Tenth ACN SIGKDD International Conference on Knowledge Discovery and Data Mining*. Presented at the tenth ACN SIGKDD International Conference on Knowledge Discovery and Data Mining. Seattle, WA. pp. 3–11. DOI:10.1145/1014052. 1014056
- [7] Lemke C, RiedelS, Gabrys B. (2009). Dynamic combination of forecasts generated by diversification procedures applied to forecasting of airline cancellations. In *IEEE Symposium on Computational Intelligence for Financial Engineering, 2009*. CIFE'09 (pp. 85–91).
- [8] Chiang, W-C, Chen, JC and Xu, X. 2007. An overview of research on revenue management: current issues and future research. *International Journal of Revenue Management*, 1(1): 97–128. DOI:10.1504/IJRM.2007.011196
- [9] Domingos, P. 2012. A few useful things to know about machine learning. *Communications of the ACM*, 55(10): 78–87. DOI:10.1145/2347736.2347755
- [10] Yoon, MG, Lee, HY and Song, YS. 2012. Linear approximation approach for a stochastic seat allocation problem with cancellation and refund policy in airlines. *Journal of Air Transport Management*, 23: 41–46. DOI:10.1016/j.jairtraman.2012.01.013



[11] Chen, T and Guestrin, C. 2016. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM. pp. 785–794. Available at <http://dl.acm.org/citation.cfm?id=2939785> [Last accessed 22 January 2017]. DOI:10.1145/2939672.2939785

[12] Kimes, SE and Wirtz, J. 2003. Has revenue management become acceptable? Findings from an International study on the perceived fairness of rate fences. *Journal of Service Research*, 6(2): 125–135. DOI:10.1177/1094670503257038