**Optimal Hotel Pricing using Machine Learning**

Megha Shah

Department of Computer and Information Science, Gannon University

GCIS 605: Scholarship Seminar

**INDEX**

1. **Abstract**
2. **Introduction**
3. **Problem Statement**
4. **Theoretical Framework**
5. **Literature Review**
6. **Research Overview**
7. **Research Questions**
8. **Research Methodology**

* **Data Source & Summary**
* **Data Preparation & Analysis**

1. **References**

**Abstract**

For revenue management in the Hotel Sector precise prediction plays a major role. Optimal pricing has long been the main purpose of revenue management systems used by the airline, rental car, and hospitality industries. To estimate hotel demand, a variety of integrated models, including time series and advance booking models, are widely utilized. Prices can be changed based on supply and demand by revenue managers. Sales can be increased through revenue management, which can also increase profitability and adjust prices to meet demand. It also provides information on customer behavior. We offer a dynamic pricing technique for managing hotel income in this study. Three components make up the dynamic pricing structure that is being proposed. The initial step entails gathering any hotel's historical data records. The examination of many factors, such as price, room availability, and client feedback, is covered in the second section to provide dynamic pricing. The final stage involves using a few classifiers to build dynamic hotel pricing using the already available data. By effectively implementing the suggested strategy to the hotel revenue management challenges, we have tested the technique. The provided models aid practitioners in improving revenue optimization and prediction accuracy, and they offer a starting point for further study into creating machine learning models for controlling hotel revenue.

**Keywords:** Dynamic Pricing, Revenue, Regression, Estimation

**Introduction**

Adjusting the price and/or quantity offered in a restricted supply to maximize profit is known as revenue management. Recently, the hotel sector has used additional revenue management systems internationally, especially at higher-rated properties. There are two types of hotel revenue schemes [Abdel Aziz et al., 2011.]. The rooms are divided into the first category, the quantity control technique, according to the length of stay, rate, room type, and guest type. Although the quantity of rooms given to each category is dynamically changed to maximize revenue, each category has a defined price. The dynamic pricing technique, on the other hand, classifies all identical rooms into a single category and uses a price that is continuously altered over time in response to variations in producers and consumers. To optimize revenue, the dynamic price is set while accounting for hotel occupancy and both current and future demand. Certain internet travel plans make use of pricing decisions. However, because of their improved forecasting capabilities, machine learning technologies are growing in prominence. Utilizing machine learning models, researchers can identify non-parametric patterns in data without adhering to rigid statistical hypotheses.

Our final objective was to create a system that would enable us to suggest dynamic hotel rates automatically based on current market information for upcoming check-ins. The suggested pricing should reflect the data-driven optimum price that the algorithm predicts will generate the greatest amount of income for each hotel. The models, however, are either unable to acquire data from the full booking arcs or are unable to understand the complex relationships between earlier bookings and actual arrivals.

Any type of organization must have a means of generating income. Of course, for your business to grow and succeed, you'll need to make enough money to cover your costs as well as some additional revenue. You will require a more advanced pricing strategy in the current competitive market, as well as the adaptability to make changes as necessary. Price optimization and revenue management are the two most crucial concerns that every organization must address. Businesses with a higher chance of surviving despite the current state of the market are those that can actively respond to market demand, build brand awareness through pricing decisions, and use inventory sensibly. Dynamic pricing enables a business to become more competitive and bounce back from poor decisions. For example, a sale on tickets for a certain flight or flights, for instance, might help an airline make up for a lack of sales at times of low demand or right before a departure.

In reaction to both exogenous and endogenous demand drivers, businesses utilize machine learning techniques to dynamically optimize their prices. Pricing methods are much more effective when using dynamic pricing. For the best pricing, it is essential to be able to precisely forecast customer demand. This is made up of two crucial parts:

1) Being able to use pricing to accommodate significant daily swings in the number of customers wanting to stay in one of the hotels in this market and understanding consumers' opinions of the relative desirability of rival hotels and their degree of price sensitivity

2) Being able to use pricing to meet consumers' assessments of the relative desirability of rival hotels and their level of price sensitivity.

Dynamic pricing alters hotel rates to raise occupancy and consequently income. By keeping an eye on market trends, evaluating your prices against those of your competitors, and setting your hotel's room rates at market value, you can increase RevPAR (revenue per available room) [Eva Lacalle et al.]. It might be a win-win situation if you adjust the pricing based on what your guests are willing to spend. The hotel is happy because it has a high occupancy rate, and guests are happy because they are paying a fair price. No of the time of year, it is also possible to handle the rates for unsold rooms smartly, potentially reducing the risk that these rooms will remain unoccupied. Increasing revenue is a result of higher occupancy.

**Problem Statement**

For our project, we established two overarching business goals, which we will further develop in the general model and analytics goals. To begin with, the main objective of the project and our main business goal is to increase annual profit (through optimization). All other goals are aimed at adjusting or limiting this essential purpose to improve the viability of our final solution. The creation of a dynamic pricing tool that determines a daily price based on variables that fluctuate and affect the price of reservations is our second business objective. Our final objective was to create a system that would enable us to suggest dynamic hotel rates automatically based on current market information for upcoming check-ins. The suggested pricing should reflect the data-driven optimum price that the algorithm predicts will generate the greatest amount of income for each hotel. When predicting hotel demand, techniques like time series, advance booking models, and other integrated models are routinely employed. Numerous parameters in the dataset have different implications for the price per room. With the methods now in use, price is dynamic and reliant on several variables.

Less than 10% of hotels already employ analytical techniques for dynamic pricing. Most of the remaining businesses simply follow their gut and change their prices seasonally [Ankit et al.]. They consider the competition as well as other market indicators when determining whether to increase or decrease the pricing. When they created their initial RM software systems at the beginning of this century, hotel chains like Marriott, Hilton, and InterContinental were among the first to employ dynamic pricing. Lodging Technology Study [Abigail et al.] in 2015, supposed that 60% of the dynamic pricing was used by United States hotels. Our final objective was to create a system that would enable us to suggest dynamic hotel rates automatically based on current market information for upcoming check-ins. The suggested pricing should reflect the data-driven optimum price that the algorithm predicts will generate the greatest amount of income for each hotel.

* Losing the hotel revenue because they are unable to capture the historical data from the entire booking curve and failed to recognize optimum forecasting of hotel demand.
* The problem statement represented in “Hotel Dynamic Pricing “is that they are unable to capture the historical data from the entire booking curve. Also, they failed to recognize the optimum forecasting of hotel demand.

Diagram

Description automatically generated**Theoretical Framework and Hypothesis:**

By taking the Categorical and Numerical Attributes for optimal Hotel pricing, we developed a Research Hypothesis.

H1: Perceived No. of reviews will positively affect customer behavior for optimal hotel pricing.

H2: Perceived Amenities will positively affect the optimal hotel pricing.

H3: Perceived Bedroom type will positively affect the optimal hotel pricing.

H4: Perceived Security Deposit will positively affect the optimal hotel pricing. Fig.1 A Proposed Model for Optimal Hotel Pricing

(Nwokeji, Joshua. C, 2022).

**Theoretical Framework and Research Questions:**

Diagram

Description automatically generatedBy taking the Categorical and Numerical attributes for optimal Hotel pricing, we developed Research Questions.

RQ 1: How does No. of reviews affect customer behavior for optimal hotel pricing.

RQ 2: Do amenities give a positive effect on optimal hotel pricing?

RQ 3: Is Bedroom type considered as a factor that affects optimal hotel pricing?

RQ 4: How does Security Deposit will affect optimal hotel pricing? (Nwokeji, Joshua. C, 2022). Fig.2: A Proposed Model for Optimal Hotel Pricing

**Literature Review**

Revenue management works best in transactions with fluctuating demand and relatively fixed, very unstable stocks (AnilBilgihanc, 2016). Enhancing the customer experience and a hotel's profitability depend on being able to predict with accuracy how many rooms will be filled on any given night. Every day, the potential vanishes since people are simultaneously generating and receiving consumer experiences, which could go unnoticed [Zeithaml et al.]. The main objective of demand-based variable pricing in revenue management [S. Choi et al.,] which is based on demand forecasting for each time ahead date, is to maximize profits. Rates can be established using occupancy estimation in the context of hotel sector revenue management. In the hotel industry, accurate occupancy forecasting is essential. A dynamic pricing model is presented by Feng and Xiao (2000), in which prices are chosen over time from a list of fixed prices. Zhao and Zheng (2000) provide a stochastic demand, non-homogeneous Poisson process, and dynamic pricing strategy for the marketing of perishable products in stock.

Guadix offers a system for making decisions that will boost hotel revenue. It supports demand forecasting and takes into account both group registrations and individual. Additionally, it uses stochastic and deterministic mathematical programming approaches. A simulator is also available that compares several room inventory control strategies to determine which is best. Due to their utilization of the unique characteristics of bookings made during the booking window, pickup models are popular in the hotel industry [Zakhary et al., 2008].

**Research Overview**

Many questions arose in the course of analyzing customer satisfaction with e-wallets. “Does hotel dynamic pricing take into account the occupation and purchasing power of the customer??”, “Is it possible for the dynamic pricing machine learning model to handle situations such as pandemics, wars, inflation, etc. without human intervention?”, “When determining optimal pricing, will demographics and recommendations of customers be considered??”.

All of these questions are based on the assumption that we can solve these issues. All of these questions revolve around the prospect of overcoming these obstacles. A comprehensive theory that captures the strategic features of pricing, such as price structures, price-demand interactions, and so on, has been developed [Warren et. al.,]. It is a known reality that for marketing to grow, the primary focus must be on customer needs, and consumer occupation has a significant impact on those needs. Customers' occupations can also influence pricing One advantage of dynamic pricing is that it can find the optimal solutions based on those attributes. Based on present capabilities, it may be possible to develop a system that can configure a dynamic pricing ml model to deal with pandemics, wars, and inflation. Without a doubt, client demographics have an impact on dynamic pricing. It could be a significant parameter influencing the pricing.

**Research Questions:**

RQ1: About the customer (P), how will the machine learning model prioritize hotel amenities (I) about other parameters (C) for optimal pricing to (O) retain the customer and make the hotel profitable(C)?

RQ2: With regards to previous customer demographics(P), how does the ML model adjust to the new demographics of the customer (I) compared to prior reports of demographics (C) for the model’s scalability (O) when the new demographic values are introduced (C)?

**Research Methodology:**

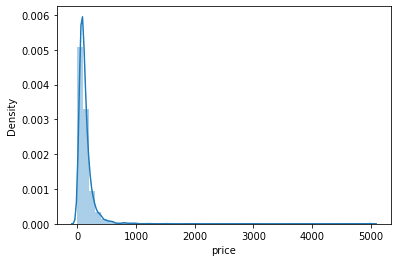
**Data source and summary**

The dataset highlights San Francisco's covered downtown location. In addition to the target attribute, there are 5207 samples and 28 other attributes. Both categorical and numerical attributes are included in it. We are using quantitative Methodology to analyze our sample size.

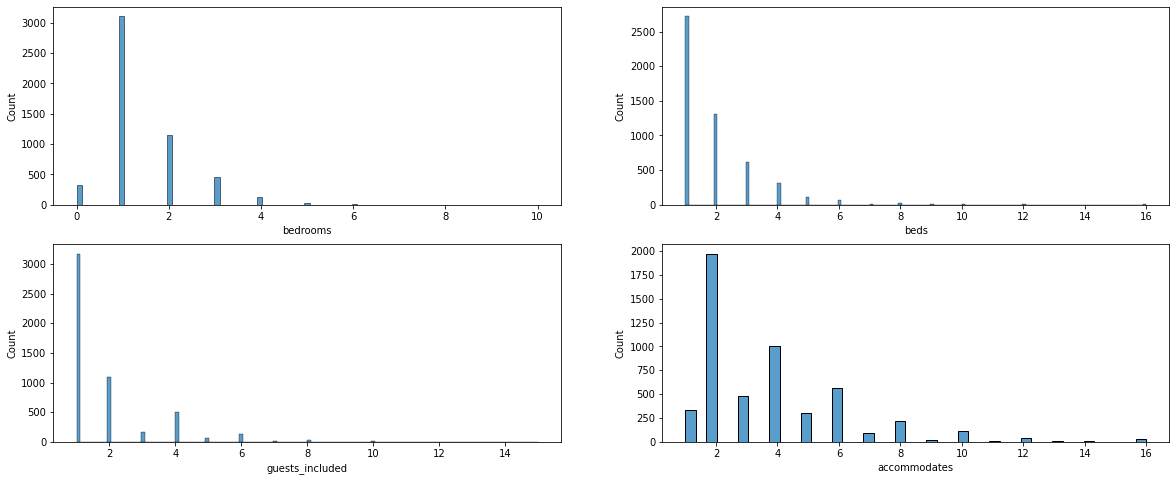
* Categorical attributes like amenities, room type, cancellation policy, etc. Whereas numerical attributes like price, maximum nights, number\_of\_reviews, etc.
* It merely displays the relative capacity, normalizing the hotel's capacity to 1 to prevent the hotels from being identified. However, the model makes use of all pertinent data, including actual capacity, which we shall demonstrate is crucial to the derivation of the best pricing rule.

**Data preparation and analysis**

* There are no blank values in the dataset.
* Using string operations, you can get rid of the dollar sign ($) and other undesirable characters like ", ',', etc. from numerical columns (such as price and cleaning charge).
* Convert lists of amenities into the count.
* Convert categorical columns such as "room type" and "cancellation policy"—into a single hot-encoded column.
* Set the dataset's mean and standard deviation to 0 and 1, respectively

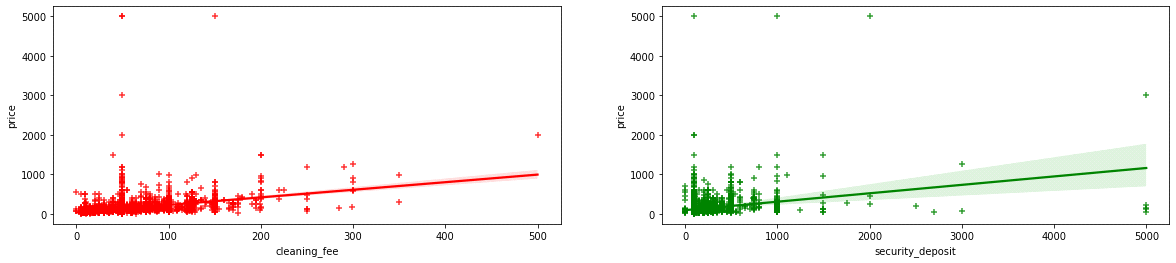


Represents the right-skewed distribution of the "price" variable, which must be transformed into a log scale before being fed into the model. The "price" factor, which can be from a few hundred dollars to a thousand dollars, relies on 27 other criteria. The fact that the mean hotel price is 136 dollars while 75% of hotels have prices that are less than or equal to 155 dollars indicates that there are many outliers with high prices due to skewness.



Given the bar, plots are designed for clients like government employees, business professionals, tourists, etc. who remain the entire weekend. We can infer from the plots provided that the majority of rooms have 1 or 2 beds. All of these graphs are skewed because this problem also affects other metrics like the number of visitors (Customers), accommodations, etc.

The cleanliness fees and room deposit fees have a significant impact on the pricing, according to the Pearson correlation, which shows a correlation of 0.6 and 0.72 for these two parameters. However, based on the data provided, it is clear that there are outliers that can also affect the model's performance. These outliers appear when the hotel's price is over the 75% quantile.



**Modelling**

The price of the target characteristic must be predicted using 58 independent attributes in this regression issue. A split of 80:20 (train-test) is used for the data set. We ended up with 1042 testing samples and 4165 training samples. I used two distinct strategies to tackle this issue:

**Method 1:**

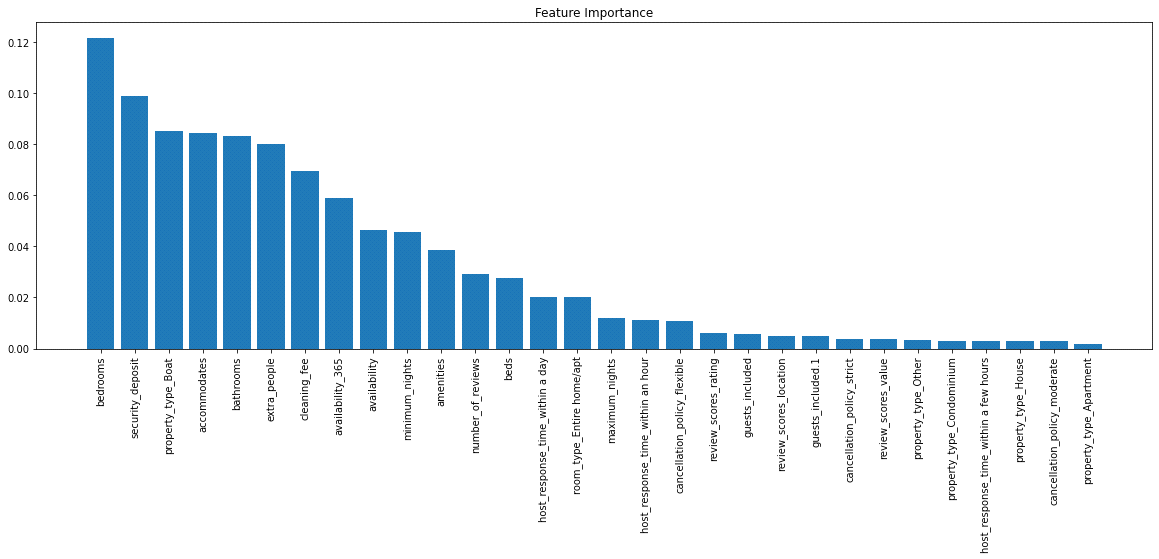
The Random Forest Regression technique employs an ensemble learning approach for regression and is a supervised learning algorithm. The ensemble learning method combines predictions from various machine learning algorithms to provide predictions that are more accurate than those from a single model to create and display the random forest models, we have used random Forest [Liaw and Wiener, 2002 et al.,] packages. Because there are too many outliers in this dataset, it is preferable to employ techniques that attempt to reduce impurity during fitting. Overfitting can be decreased through pruning. I used the hyperparameter tuning technique to determine the ideal random forest parameters before fitting the model.

The best parameters of random forests are:

max depth - 11

Number of estimators - 120

Criterion – log



Following training on the ideal features, the random forest model generated the importance of the features in descending order.

**Method 2:**

Multi-Layer Perceptron Regression (MLP): The Activation function is the Number of layers perceptron, which is implemented by using this method, this technique has no activation function in the O/P layer which is used to train using backpropagation.It is a neural network that inputs normalized data, performs some preliminary processing, and outputs the outcome. It's a type of "black box" model that doesn't need any manual engineering, unlike a random forest, however training these models to outperform conventional machine learning models requires more data. It is prone to anomalies. Using hyperparameter tuning we obtained the following optimal parameters:

Hidden layer size - 34

Solver - adam

Alpha - 0.0005

Activation - relu

Number of iterations – 7000

**Method 3:**

CatBoost expands on gradient boosting and decision tree theory. The basic goal of boosting is to successively combine a large number of weak models or models that just slightly outperform chance, in order to produce a strong, competitive predictive model through greedy search.

By training a series of weak learners, the boosting algorithm's main benefit is its ability to handle outliers. In contrast to previous methods, it also works well on samples outside of distribution. Using hyperparameter tuning we obtained the following optimal parameters:

Learning rate - 0.03

depth - 8

L2\_ref\_reg - 0.2

Number of iterations – 150

**Result:**

Table

Description automatically generated

The same number of samples and data distribution were used for both model training and testing. With a mean absolute error of 25.39 on the train set and 43.93 on the test set, the CatBoost approach performed better than the neural network and random forest model. Even while CatBoost performed well, unlike the MLP regressor, it is not extended to the test set, which causes overfitting. The MLP regressor will perform better if the dataset has more samples because neural networks are data hungry.

**Conclusion:**

We describe a dynamic pricing method for hotel revenue management in this research. The proposed dynamic pricing system is composed of three components. The initial step is to gather any hotel's historical data records. In the second section, numerous characteristics such as price, room availability, and customer reviews are analyzed in order to provide dynamic pricing. The final stage entails running a few regression algorithms on the supplied data to generate dynamic hotel pricing. We proposed a novel strategy that used regression algorithms. The Random Forest Regressor, catBoost and Multilayer Perceptron Regressor algorithms were used, and the results were recorded. The random forest strategy outperformed the neural network model. Even though the random forest outperformed the MLP regressor, it was not extended to the test set, resulting in overfitting. Because neural networks are data-hungry, the MLP regressor will perform better if the dataset contains more samples.

**References**

Abdel Aziz, H., Saleh, M., Rasmy, M. and El-Shishiny, H. (2011) Dynamic room pricing model for hotel revenue management systems. *Egyptian Informatics Journal*. 12(3):177-183.

Abigail A. Lorden, (2016). *Lodging Technology Study,* https://hospitalitytech.com/2016-lodging-technology-study

Ankit Raj, (2019) *Hotel Dynamic Pricing*, https://tech.goibibo.com/hotel-dynamic-pricing-7042a561f8a3

Eva Lacalle, (26 Apr 2021) *How can dynamic pricing improve your hotel revenue?* www.mews.com/en/blog/dynamic-pricing-hotels

G. Cetin, T. Demirciftci, and A. Bilgihan. Meeting revenue management challenges: Knowledge, skills and abilities, *International Journal of Hospitality Management*, vol. 57, [Online] pp. 132 – 142, 2016. (AnilBilgihanc, 2016)

Ingold, A., McMahon-Beattie, U. and Yeoman, I. (2000) *Yield Management: Strategies for the service industries*, (pp. 233-255). London: Continuum.

J. Nick and L. Darren, (1996) Hotel revenue management and its impact on customers’ perceptions of fairness, *International Journal of Contemporary Hospitality Management*, vol.8, no.2, pp. 14–16. https://doi.org/10.1108/09596119610111677

Liaw, A. and Wiener, M. (2002). *Classification and regression by random forest*. R News, 2(3):18–22. https://cogns.northwestern.edu/cbmg/LiawAndWiener2002.pdf

Qing Zhang, Liyuan Qiu, Huaiwen Wu, Jinshan Wang, Hengliang Luo, (2019) Deep Learning based Dynamic Pricing Model for Hotel Revenue Management, *International Conference on Data Mining Workshops (ICDMW),*

https://www.researchgate.net/publication/338602035\_Deep\_Learning\_Based\_Dynamic\_Pricing\_Model\_for\_Hotel\_Revenue\_Management

R. N. Warren, (2017) *Occupancy forecasting methods and the use of expert judgement in hotel revenue management*. https://doi.org/10.31274/etd-180810-5867

S. Choi and A. S. Mattila, (2004) Hotel revenue management and its impact on customers perceptions of fairness, *Journal of Revenue and Pricing Management*, vol. 2, [Online], no. 303, pp. 303–314.https://doi.org/10.1057/palgrave.rpm.5170079

Talluri, K.T. and Van Ryzin, G. (2005), The theory and practice of revenue management, *Journal of revenue and pricing management,* http://dx.doi.org/10.1057/palgrave.rpm.5170123*.*

V. A. Zeithaml, A. Parasuraman, and L. L. Berry, Problems, and strategies in services marketing, *JournalofMarketing*, vol.49, no.2, pp. 33–46,1985. <http://www.jstor.org/stable/1251563>

Yao, Z., Xu, X., and Yu, H. (2018). Floor Heating Customer Prediction Model Based on Random Forest. Proceedings - 17th IEEE/ACIS *International Conference on Computer and Information Science*, ICIS 2018, pages 573–578.

https://ieeexplore.ieee.org/document/8466420

Yueqin Zhang, (August 2019), the Graduate School of Cornell University, *forecasting hotel demand using machine learning approaches* https://www.researchgate.net/publication/340134067\_Forecasting\_Hotel\_Demand\_Using\_Machine\_Learning\_Approaches

Zakhary, A., El Gayar, N., and Atiya, A. F. (2008). A comparative study of the pickup method and its variations using simulated hotel reservation data. *ICGST international journal on artificial intelligence and machine learning*, vol.8, pp.15–21.

Zhao, W.and Zheng, Y.S. (2000) *Optimal dynamic pricing for perishable assets with nonhomogeneous demand*. Management Science. Vol.46(3), pp.375-388.

Nwokeji, Joshua. C (2022). *Theoretical Framework and Research Methodologies* [Lecture 6, slide7, 8].