Optimizing Q-Learning for Dynamic Pricing in Airbnb

Chitra M[1], S Sangeetha[2], Anshidha Anish[3] and Sruthika R[4]
{chitram2@srmist.edu.in[1], ss0200@srmist.edu.in[2], aa0784@srmist.edu.in[3], rr7269@srmist.edu.in[4]}

Department of Computer Science Engineering, SRM Institute of Science and Technology, Chennai, Tamil Nadu, India[1, 2, 3, 4]

Abstract. Dynamic pricing is the lifeblood of revenue generation in online marketplaces like Airbnb. Reinforcement learning (RL), in particular Q-learning has been extensively investigated in this problem, but the problem becomes challenging since the state space is huge and the learning dynamics are unstable. This work contributes with a novel, multi-agent Q learning, optimization framework for dynamic pricing that additionally combines adaptive open loop exploration-exploitation strategies and economic based reward functions. To show the efficiency of the proposed model we compare with a simple rule-based pricing rule. Experimental results indicate that our Q-learning model generates higher revenues as well as higher dwelling yield with a higher pricing comparatively to heuristics, the difference between revenues of two methods being statistically significant.

Keywords: Dynamic Pricing, Q-Learning, Airbnb, Multi- Agent Systems, Reinforcement Learning

1 Introduction

Dynamic pricing is now a key tool for revenue optimization in some industries, we can mention for example the hospitality, e- commerce, and transportation ones. In context of the short-term rental market, platforms such as Airbnb also offer hosts with price-setting mechanisms and many of them take their pricing decisions based on the demand dynamics, competitive pressure and seasonal trends [1]. Conventional pricing methodologies often based themselves on fixed rules, historical means or on manually designed pricing strategies which may resistant to external market movements. These constraints stress the requirement of more flexible and data-driven pricing policies.

Reinforcement learning (RL) is a promising approach to address dynamic pricing problem by learning optimal policy by interacting with environment [5]. In the RL models, the Q learning RL model is model-free approach in which agents learn a pricing policy for information goods without the explicit modelling of competition [6]. However, it is much more challenging to Apply Q-learning to dynamic pricing, which suffers from the issues of the high-dimensionality of state space, the instability of learning convergence and the susceptibility to market dynamics. These issues could in the end lead to a negative impact on pricing decisions and corresponding revenue [2].

In this work, we develop a novel efficient multi-agent Q-learning framework for Airbnb dynamic pricing. We contribute to advancing the state of learning by stabilizing learning, reducing price volatility for learning, and enhancing decision making [3]. 1 We study how

ICITSM-Part I 2025, April 28-29, Tiruchengode, India

DOI 10.4108/eai.28-4-2025.2357857

reinforcement learning can support human-designed pricing policies to improve revenue management in dynamic rental markets [4].

We thus contribute to machine learning in dynamic pricing demand by introducing a Q-learning application in the Airbnb market. In particular, the implication of this work is that we can designed more power RL based models for pricing that are adaptive to the dynamic nature of the market and formulate the optimized pricing and admission strategies for the hosts, with higher added value than the others pre-existing solutions [1].

2 Literature Review

Dynamic pricing strategies are used in many industries (e.g., e-commerce and hospitality) to stay competitive and react to supply and demand changes in real-time. In this section, we review the dynamic pricing literature, differentiating between traditional operations research approaches and more recent reinforcement learning methods.

Historically, dynamic pricing strategies have relied on deterministic and stochastic optimization models. For instance, businesses have implemented statistical models to predict demand elasticity and adjust prices accordingly. Traditional rule-based approaches often incorporate predefined heuristics and regression-based techniques to estimate optimal price points. Agnihotri and Raj [1] proposed a Bayesian Q-learning algorithm for dynamic pricing, enhancing automated pricing in electronic marketplaces by utilizing past observations. How- ever, while these models offer structured pricing frameworks, they struggle to respond effectively to sudden market fluctuations and evolving consumer behavior.

Traditional pricing systems are typically heavily human involved with a set of fixed automation rules, which are inflexible and hopelessly inefficient, i.e., cannot modify patterns in a timely manner, to react to the market. Static statistical models are not able to learn from new real-time data and the resulting prices of the product are incoherent. As noted by Shen et al. [2], classical procedures may not be adapted to dynamic e-commerce markets, where price decisions have to be constantly updated according to competitive pressures and demand co-evolution.

Dynamic pricing Reinforcement learning (RL) has been proposed as a potential alternative for dynamic pricing where it can learn an optimal pricing strategy through continual interaction with the market. Unlike standard methods, RL based frameworks can change prices in a dynamic way using an online feedback, thus enabling the businesses to learn and maximize revenue and support the equilibrium in the economy.

Recent researches have investigated the use of a number of machine learning algorithms for dynamic pricing. Mahyoub et al. [3] investigated Airbnb pricing pre- diction with ma- chine learning models and demonstrated that predictive algorithms assist hosts in determining competitive prices. Likewise, Singh [4] arrived at a reinforcement learning solution adapted to the hospitality industry, which reacts to the market movements adjusting the pricing strategies. These articles use machine learning to suggest how demand-responsive automated pricing changes could react faster to demand shifts. Shen et al. [2] proposed a deep reinforcement learning framework for e-commerce markets, which employs sophisticated models to make better pricing decisions. Upadhyay et al. [3] analyzed conventional Q-learning and SARSA based pricing models and showed that the deep Q-learning with an annealing process enhanced pricing efficiency. These results are consistent with that observed by Han et al. [5], which proposed that multi- agent RL frameworks can improve upon existing price optimization techniques through adaptive exploration-exploitation strategies, and market-aware reward functions.

With the rapid development of reinforcement learning, we believe that future re- search will put more emphasis on joint-optimization models, which are models that combine traditional optimization methods with deep learning. Mustafina et al. [3], which emphasized the necessity of gene algorithms and Bayesian optimization used to optimize the hyperparameters of reinforcement learning pitted (further bettering the pricing strategy). The increasing use of AI-based pricing models reflects a move to increasingly autonomous and adaptable decision making in dynamic pricing arenas.

3 Methodology

This study presents a multi-agent Q-learning framework designed to optimize dynamic pricing strategies for Airbnb rentals. The methodology is structured to ensure an adaptive pricing model that balances revenue maximization with market competitiveness. The structure consists of several key elements, preprocessing, including reinforcement data extraction, learning model training, evaluation, and continuous adaptation. As shown in Fig 1, the system consists of four main com- ponentes: Data Collection, Feature Engineering, Multi-Agent Q-Learning, and Price Optimization. Historical Airbnb data is transformed into meaningful features that inform dynamic pricing strategies. The Q-learning model enables us to simulate hosts that constantly learn from the dynamic of the market, making the optimal price decision. This repeating process also guarantees that the resulting price suggestions are both revenues optimized and competitive in terms of occupancy.

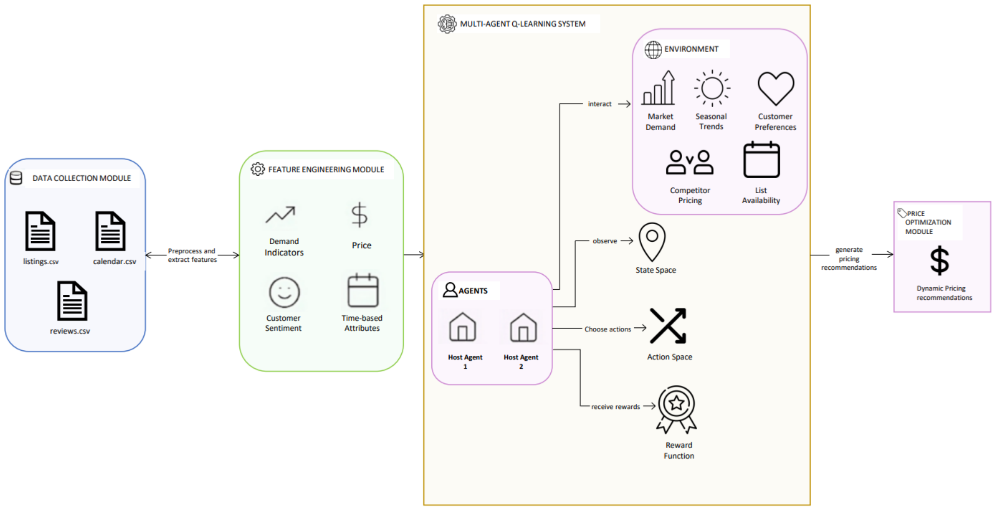

Fig. 1. System Architecture of the proposed model.

3.1 Model Design

The proposed model adopts an organized pipeline to realize a well-refined multi-agent Q-learning method for dynamic pricing. The system naturally receives input data from different channels, leverage reinforcement learning algorithms to find the optimal price policy, and has a real-time feature update according to market variations. The subsets of construction components have been added incrementally and refined, to provide robustness and security to the architecture model in practical situations.

3.2 Data Preparation

The data preparation phase is crucial in structuring the model's inputs to ensure optimal learning outcomes. The steps involved in this phase, corresponding to Algorithm are outlined as follows:

- Feature Engineering: Chooses important features like shifts in demand, competition pricing, room occupancy, seasonality and customer behavior.
- Normalization & Scaling: Decreases variance when looking at numerical features so that the performance of the model increases in consistency.
- Market Segmentation: Categorizes rental properties based on location, property type, and amenities to allow customized pricing strategies.
- Action Space Definition: Defines possible price adjustments that the RL model can choose from, ensuring smooth price variations.
- State Representation: Encodes the market environment at any given moment using dynamic state variables such as demand elasticity and supply saturation.
- Reward Function Design: Incentivizes the model to optimize both revenue generation and occupancy rates while maintaining competitive pricing.
- Training-Testing Split: Partitions historical data between training and validation sets to determine the generalization capacity of a model.

3.3 Multi-Agent Q-Learning for Dynamic Pricing

The reinforcement learning system functions within a multi- agent Markov Decision Process (MDP) framework, where several independent agents each represents an Airbnb host engage with a common market environment to develop and refine effective pricing strategies. Each agent in the framework acts as an individual Airbnb host, dynamically adjusting its pricing to optimize revenue while simultaneously competing with other agents in the shared market environment. The learning process follows the Q-learning update rule, refining price optimization based on observed rewards:

$$Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma \max_{a'} Q(s',a') - Q(s,a)] \tag{1}$$

- $Q\ (s, a)$ represents the expected reward for acting a in state$s$.
- α is the learning rate, controlling the speed of updates.
- γ is the discount factor, balancing short-term and long- term rewards.
- r is the reward obtained from revenue and occupancy.

3.4 Multi-Agent Interaction and Reward Mechanism

Unlike single-agent Q-learning, multi-agent Q-learning con- siders the interdependence between different pricing strategies. Each agent observes market dynamics, adjusts pricing decisions, and earns from competitive interactions.

The reward function is designed to incentivize revenue maximization while discouraging excessive underpricing or price instability:

$$Rt = \alpha \times \mathbf{R} - \beta \times \mathbf{V} + \gamma \times \mathbf{C} \tag{2}$$

where:

- R: Revenue, which is the booking income.
- V: Vacancy Penalty = discourages extreme pricing, causing too often un-booked nights.
- C: Competitive Pricing Adjustment, the type of price adjustment to follow the market expectations and keep margins.
- Each agent resorts its pricing policy through combine of:
- Exploration: Testing new prices.
- Exploitation: Leveraging learned optimal prices.
- Opponent-Aware Learning (OAL): Focuses on how to adapt the bidding strategy according to their opponent's pricing behaviour.

3.5 Model Training and Optimization

The reinforcement learning model is trained using historical Airbnb data, simulating thousands of pricing decisions over multiple time steps to learn an optimal pricing policy. The training process contains the following steps:

- Multi-Agent Exploration-Exploitation Balance: epsilon-greedy method used for each agent to explore new price strategies while exploiting learned optimal prices in the presence of competitive interactions.
- Opponent-Aware Q-Learning: Each agent modifies its pricing behaviour dependent on observed competitor strategies, providing a means for dynamic price adaptation to market fluctuations.
- Adaptive Learning Rate Scheduling: Adjusts learning rates on the fly for each agent to avoid premature con- vergence and allow an overall smooth adaptation to new demand patterns.
- Multi-Agent Deep Q-Networks (DQN) Integration: The difference with traditional q-learning is how deep q networks use the deep neural network that acts as a function approximator of Q-values, making it easier to be able to pick an action from a set of large state spaces.
- Bayesian-Genetic Hyperparameter Optimization: Harnessing the power of both Bayesian optimization and genetic algorithms, this technique hyper parametrizes several crucial RL parameters such as discount factors, learning rates and exploration strategies to ensure faster convergence and better performance.

- Market Feedback and Continuous Adaptation: Builds in a real-time feedback loop to evolve pricing policies based on recent booking outcomes, clamorous seasonality variations, as well as changes in consumer behaviour keeping prices effective.

3.6 Model Evaluation and Baseline Comparison

To assess the effectiveness of the multi-agent Q-learning model, it is tested against baseline pricing strategies commonly used in Airbnb markets. These include:

- Fixed Pricing Model: A permanent pricing method in which rental rates do not change over time, because they are based on the historical trends of demand.
- Rule-Based Dynamic Pricing: Price adjustment based on certain heuristics defined by the rules such as increases in price during peak periods and decreases in low demand time.

3.7 Evaluation Metrics

Key Performance Metrics The performance of the proposed multi-agent Q-learning model is assessed with the help of the following critical metrics:

a. Revenue Maximization: Evaluate the total revenue gains generated after the trial period in order to infer profit gains.
b. Occupancy Rate: Measures the ratio of booked nights versus total available nights, which represents the reaction of demand.
c. Pricing Stability: Considers the difference in the price adjustments over time to avoid the extremely high price actuations.
d. Market response: Measures how well the model adapts to changes in demand, price from the competitor, and seasonality.
e. Convergence of learning: Studies the number of iterations the Q-learning agents take to stabilize their pricing strategies, achieving learning efficiency.

4 Results and Discussion

This section analyses the performance of dynamic pricing model that is built on multi-agent Q-learning and compares it with a rule-based pricing approach. This is based on the KPIs such as total revenue and occupancy rate.

4.1 Baseline vs. Multi-Agent Q-Learning Performance

Starting with a simple rule-based pricing strategy (where prices are based on basic demand heuristics) to benchmark performance of the multi-agent Q-learning model We compared it with our optimized RL based model, which adjusts pricing based on the conditions of the market and learns from past interactions.

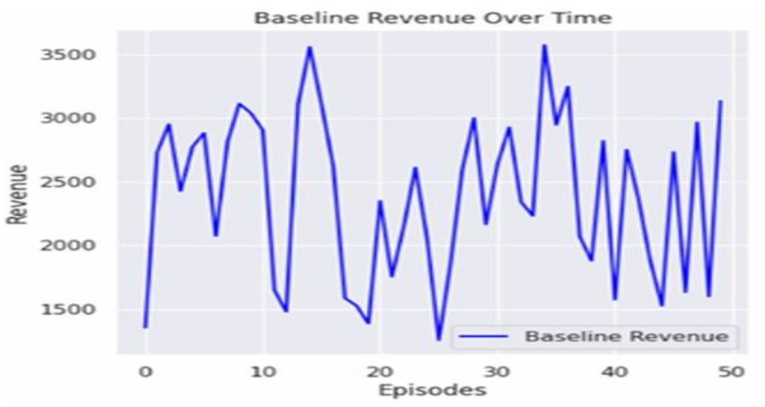

Fig. 2. Baseline Revenue Over Time.

Revenue Comparison: Fig 2 presents the revenue trends across training episodes for both the baseline model and the multi-agent Q-learning model. The baseline approach results in fluctuating revenue with no significant improvements. In contrast, the RL model exhibits a clear increasing revenue trend, demonstrating its ability to learn and optimize pricing strategies for better profitability.

Fig. 3. Revenue Smoothing.

Revenue Smoothing: Fig 3 illustrates the revenue progression of the optimized multi-agent Q-learning model after applying smoothing techniques. Unlike the initial RL model, which exhibited instability in price adjustments, the refined approach stabilizes pricing decisions, leading to consistent revenue generation and reduced variance.

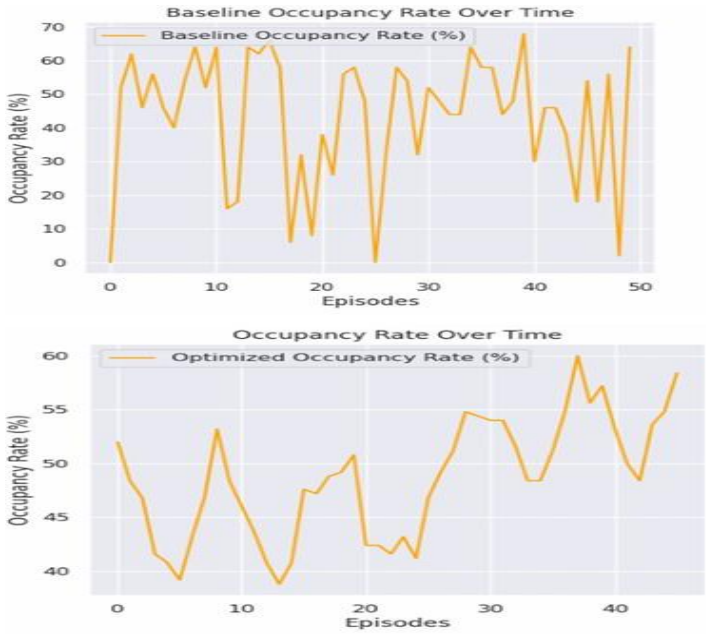

Fig. 4. Comparison of baseline model and multi-agent Q-learning approach.

Occupancy Rate Comparison: Fig 4 compares the occupancy rates for both the baseline model and the multi- agent Q-learning approach. The rule-based strategy results in lower and more inconsistent occupancy rates, whereas the Q- learning model achieves higher and more stable occupancy levels, indicating improved market adaptability.

4.2 Quantitative Comparison

To further illustrate the performance improvements, Table 1 presents comparative analysis of key performance metrics between the baseline pricing strategy and the multi-agent Q- learning approach.

Table 1. Quantitative Comparison.

Metric	Baseline Pricing	Optimized Multi-Agent Q-Learning
Average Revenue per Episode	$2,438.07	$3,512.72
Final Revenue (Last Episode)	$1,853.71	$2,958.55
Average Occupancy Rate	44.44%	48.68%

5 Discussion

The experimental studies show our multi-agent Q-learning model can achieve pricing efficiency substantially better than a basic rule-based pricing strategy. The RL pricing model adaptively modifies the prices in response to the changes of the demand and the behaviors of the competitor, achieving better revenue and occupancy. The multi-agent learning allows for the agents (representing various property offers) to interact and respond to competitive pricing forces resulting in an increased robustness against a dynamic market. Also, the stabilized- Q-learning naturally exhibits less variability in the pricing decision, which result in trusty market.

6 Conclusion

This study demonstrates the benefits of employing an MA Q-learning approach to dynamic pricing in the Airbnb market. In contrast to the standard rule-based trading model, the proposed trading system can achieve substantial incremental revenue potential and maintain higher occupancy rates by dynamically responding to market characteristics. The multi-agent learning method of learning the optimal pricing from competitive pressure produces robust and adaptive pricing model. The stabilization of the learning also prevents undue price vacillations, and results in stable and reliable pricing policies. These results further suggest the promise of reinforcement learning for short-term rental pricing optimization and call for future investigations using deep reinforcement learning as means of achieving improved market adaptability.

References

[1] W. Han, "A dynamic pricing algorithm by Bayesian Q-learning," *IEEE*, 2010.

[2] J. Liu et al., "Dynamic pricing on e-commerce platform with deep reinforcement learning: A field experiment," *IEEE*, 2019.

[3] W. Han, L. Liu, and H. Zheng, "Dynamic pricing by multiagent reinforcement learning," *IEEE*, 2008.

[4] A. Agnihotri, "Advanced deep reinforcement learning framework for dynamic pricing optimization in e-commerce marketplaces," *IEEE*, 2022.

[5] J. Shen, Y. Wang, and F. Xiao, "Dynamic pricing strategy for data prod- uct through deep reinforcement learning," *IEEE Journals and Magazine*, 2023.

[6] F. Fitrianingsih, D. A. Rahayu, and F. R. Zazila, "Dynamic pricing analytic of Airbnb Amsterdam using k-means clustering," *IEEE*, 2023.

[7] I. Singh, "Dynamic pricing using reinforcement learning in hospitality industry," *IEEE*, 2023.

[8] C. V. L. Raju, Y. Narahari, and K. Ravikumar, "Reinforcement learning applications in dynamic pricing of retail markets," *IEEE*, 2002.

[9] J. Wang and Z. Lei, "Application of reinforcement learning in dynamic pricing algorithms," *IEEE*, 2009.

[10] M. Mahyoub, A. A. Ataby, Y. Upadhyay, and J. Mustafina, "AIRBNB price prediction using machine learning," *IEEE*, 2025.

[11] S. Chen et al., "Multi-agent Q-learning with dynamic pricing in smart grids," *IEEE Transactions on Smart Grid*, vol. 10, no. 5, 2019.

[12] T. T. Nguyen, N. D. Nguyen, and S. Nahavandi, "Deep reinforcement learning for multiagent systems: A review of challenges, solutions, and applications," *IEEE Transactions on Cybernetics*, vol. 50, no. 9, 2020.

[13] M. Abolhasani and F. Moazeni, "Dynamic pricing in cloud services using reinforcement learning," *IEEE Access*, vol. 8, pp. 115507–115517, 2020.

[14] Z. Chen, "Competitive pricing in e-commerce using multi-agent Q- learning," *IEEE Access*, vol. 7, pp. 143111–143123, 2019.

[15] Y. Zhang, J. Wang, and X. Li, "Reinforcement learning-based framework for airline dynamic pricing," *IEEE Access*, vol. 7, pp. 91246–91256, 2019.