IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 4, April 2026

DOI: 10.17148/IJIREEICE.2026.14419

A Q-Learning Based Dynamic Pricing Model for Minimizing Food Waste in Supermarkets

D. Vimal Kumar[1], A. Hemalatha[2], S. Harish[3], M. Mohammed Yunush[4]

Department of Computer Science, Rathinam College of Arts and Science Coimbatore-641021[1-4]

Abstract: Food waste in supermarkets is a big problem for the economy and the environment, especially when it comes to perishable goods that don't last long. Traditional static pricing strategies don't take into account changes in demand or the freshness of products, which can lead to unsold inventory and more waste. This study proposes a Q-learning-based dynamic pricing model that modifies product prices as they near expiration to tackle this issue.The pricing issue is framed as a Markov Decision Process (MDP), wherein the system acquires optimal pricing strategies via ongoing engagement with the environment. Important things like how long the item will last, how much demand there is for it, and how much stock there is are all taken into account when making decisions. The model's goal is to make as much money as possible while keeping unsold stock and food waste to a minimum.Experimental results show that the suggested method works better than traditional pricing methods to cut down on food waste and increase overall profits. This study shows that reinforcement learning methods can be used to create smart and long-lasting pricing systems for stores.

Keywords: Dynamic pricing, Q-learning, reinforcement learning, reducing food waste from perishable goods, the Markov decision process (MDP), and managing inventory

I. INTRODUCTION

Experimental Design The Q-learning-based dynamic pricing scheme was implemented in a simulation study by developing a virtual environment for the supermarket. Experiments were conducted through various episodes, which is defined as the lifecycle of a product from procurement to the expiry period. Some of the important factors that were considered during the testing phase are the value of the learning rate, which is denoted by 'α' with a value of 0.1, discount rate 'γ' with a value of 0.9, and using an 'ε-greedy' policy with a gradually decreasing 'ε' value. The actions considered during this study were the discounts offered, which were 0%, 10%, 20%, and 30%. use, and more use of natural resources like water, energy and agricultural inputs used during production static pricing agricultural inputs used during production static pricing systems commonly employed in supermarkets are based upon the idea that prices for all products are set once and will stay the same until modified manually by the store managers. Even though such systems are relatively simple to introduce and use, they lack dynamics inherent in real-world situations and do not take into account changing demands and perishability of products. Namely, the closer products come to their expiry date, the lower their value; nonetheless, due to the static system of pricing, this value is not taken into account. This results in consumers rejecting items close to expiry, which makes them waste, while some retailers try dealing with the problem by using manually controlled discounts, which are generally based upon rules and vary greatly from one retailer to another.

Retailers have started using dynamic pricing to get around these issues. In simple terms, dynamic pricing means adjusting prices based on what's happening in real time—things like how many people want the product, how much isleft in stock, how close it is to expiring, and how customers usually buy. When a product is getting close to its expiration date, dropping the price can help clear out inventory and cut down on waste.But coming up with the perfect dynamic pricing plan isn't easy.

There's a lot of unpredictability—customers act differently than you expect, demand goes up and down, and you have to balance making money with not throwing things away. Because these goals can clash, you can't just set a simple rule and expect it to work every time. It's a messy problem, and it keeps retailers on their toes.Machine learning has really taken off in retail and supply chain work lately. People use supervised learning models all the time for things like forecasting demand, predicting sales, and breaking customers into segments. But here's the thing: those models depend a lot on old, labeled data, and they're not great at handling decisions that play out over time in changing situations. Take supermarket pricing, for example. It's not just about making a single prediction and walking away. Every time you change a price, it affects what happens next—future demand, how much stock you have, and what kind of revenue you pull in. So you need something more flexible, something that learns and adapts as it goes. A feedback-driven approach fits the bill.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 4, April 2026

DOI: 10.17148/IJIREEICE.2026.14419

Reinforcement learning (RL), a subfield of machine learning, provides a powerful framework for solving sequential decision-making problems under uncertainty. In reinforcement learning, an agent interacts with an environment and learns optimal actions by receiving feedback in the form of rewards or penalties. Unlike traditional supervised learning methods, RL does not require labeled input-output pairs and instead learns optimal policies through trial and error. Among various reinforcement learning algorithms, Q-learning is particularly suitable for such applications due to its model-free nature, meaning it does not require prior knowledge of the environment's dynamics. This makes it highly applicable to real-world systems such as retail pricing, where demand patterns and customer responses are complex and not explicitly known.

In this study, we look at how to set prices for perishable supermarket goods using a Markov Decision Process (MDP). Here, the system keeps track of things like how many days products have left on the shelf, how much is in stock, and what demand looks like. The agent (think of it as the decision-maker) chooses prices—like adding a discount or tweaking the cost—and then gets rewarded based on two things: how much revenue sales bring in and how much waste gets cut down. As time goes on, the Q-learning agent figures out the best pricing moves to bring in steady profits while also keeping waste to a minimum.

There are many benefits associated with this suggested method, which is different from the conventional way of pricing goods in a supermarket. To begin with, the method incorporates automation in determining prices, hence less reliance on human judgment. Secondly, since the process is dynamic, it adjusts automatically to changes in its environment through learning. Thirdly, the method is scalable and may be applied to various types of products and even retail formats. Lastly, since it reduces wastage of goods, it promotes environmental conservation practices.

In addition, the availability of retail data provided by electronic invoicing, inventory management, and consumer analysis software has enabled the development of intelligent pricing. Nonetheless, although such an advance in technology has been achieved, most of the current supermarket systems utilize traditional pricing algorithms. It is evident that there is a considerable difference between the potential for implementing intelligent pricing and its actual implementation.

Another significant dimension that comes into play in this problem area is that of the customer response to changes in prices. Apart from being affected by price discounts, the customers' purchase behavior is impacted by perceptions related to product quality, immediacy of usage, and freshness. Consequently, there is a need for a pricing approach which incorporates economic considerations as well as psychological elements. In this regard, reinforcement learning stands out as the most suitable technique owing to its non-humanistic nature. Learning stands out as the most

learning stands out as the most suitable technique owing to its non-humanistic nature. Although substantial progress has been made in intelligent pricing techniques and reinforcement learning approaches, there remains an inadequate level of research conducted in the domain, where existing research is restricted to simulations based on a certain set of assumptions that do not consider real-world factors. Issues such as scalability, flexibility, and software integration remain unresolved.

A key goal of this study is to develop and test a dynamic pricing model using reinforcement learning techniques, specifically Q-learning, to minimize food waste without compromising profits. This study will prove that reinforcement learning methods offer a promising approach and can be used to supplement existing pricing models. The feasibility of this model will be demonstrated by conducting simulation experiments and comparing its performance against other pricing models.

In summary, this study makes a contribution to the emerging research domain of intelligent retail systems through its use of reinforcement learning techniques to solve practical issues related to sustainability. This research shows that reinforcement learning can be used to develop a systematic approach to incorporating artificial intelligence within supermarket pricing schemes.

II. LITERATURE REVIEW

Food waste has emerged as an extremely important issue in today's world due to its profound impact on the economy, environment, and society. Among several industries that contribute to food wastage, supermarket stores play a major role because of the type of product they deal with, which includes fresh produce like fruits, vegetables, milk products, and prepared foods. Such products do not last long and are extremely vulnerable to environmental conditions and changes in consumer preferences. Hence, many items get wasted without being sold before their expiry date. Apart from resulting in significant economic losses for the retailing firms, food waste has severe repercussions on the environment.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 4, April 2026

DOI: 10.17148/IJIREEICE.2026.14419

The conventional pricing policies in retail follow a static approach wherein the prices of goods do not fluctuate even when there are variations in the demand for the goods or their freshness levels. While the application of such pricing systems is relatively easy, they overlook the dynamic aspects of business operations in the real world. In many cases, products that have reached their expiration dates are sold at the same price as new products. As a result, customers perceive them as being less valuable and refrain from buying them. Although many retailers adopt manual pricing systems, they usually use rule-based techniques which are incapable of adapting to changing situations.

In recent years, dynamic pricing has emerged as a promising solution to address the limitations of static pricing models. Dynamic pricing involves adjusting product prices based on various factors such as demand, inventory levels, and remaining shelf life. By offering timely discounts on products nearing expiration, retailers can incentivize customers to purchase them earlier, thereby reducing waste and improving inventory turnover. However, determining the optimal pricing strategy is a complex task due to the uncertainty in consumer behavior, variability in demand patterns, and the need to balance multiple objectives such as profit maximization and waste minimization.

In order to deal with the complexities effectively, there is a need for more sophisticated computation methods. Machine learning approaches have increasingly received attention due to their ability to process large amounts of information and make intelligent decisions based on that information. Supervised machine learning algorithms have been frequently utilized for applications such as demand forecasting and predicting consumer behaviours. One of the main shortcomings of supervised learning algorithms is that they depend on past information.

The field of machine learning contains the area known as reinforcement learning, which is useful for handling sequential decision-making tasks with uncertainty. Contrary to other approaches, reinforcement learning allows the learner to acquire optimal behaviors while interacting with the environment through feedback in terms of rewards. Out of many reinforcement learning techniques, Q-learning is one of those that are particularly useful because it is model-free, hence allowing the algorithm to discover optimal policies irrespective of knowledge about environment dynamics.

For the case study under analysis, the issue of pricing perishable products within the supermarket context is defined as a Markov Decision Process (MDP). It takes into account some relevant parameters including product shelf life, customer demands, and inventory level. The Q-learning algorithm acts as an action taker and performs different pricing activities. Specifically, it sets discount rates. At the same time, it obtains rewards depending on sales volumes and waste reductions. In other words, the Q-learning algorithm learns how to achieve maximum gains through its actions.

The proposed approach offers several advantages over traditional pricing methods. It enables automated and adaptive decision-making, reducing the need for manual intervention. The model continuously improves its performance by learning from interactions, making it robust to changing market conditions. Additionally, the system is scalable and can be applied across different product categories and retail environments. By reducing food waste, the approach also contributes to environmental sustainability and aligns with global initiatives promoting responsible consumption and production.

The objective of this study is to design and test a novel Q-learning based dynamic pricing strategy which will overcome the constraints of the existing models. The performance of the proposed model is tested using simulation techniques and a comparison is made between the proposed model and other pricing strategies such as static and rule based pricing.
In summary, this research adds to the existing literature on the combination of artificial intelligence and retail management. The study offers a viable solution to reduce food wastage without sacrificing profit, which is an important issue not only economically but environmentally.

III. PROPOSED METHODOLGY

This research proposes a Q-learning–based dynamic pricing framework designed to minimize food waste in supermarkets while maintaining profitability. The methodology is structured into multiple stages, including problem formulation, data modeling, system design, learning process, and evaluation.

A. Problem Formulation

The dynamic pricing model for perishing products is designed using the Markov Decision Process. Here, the dynamic pricing model can be viewed as an agent that interacts with its environment to determine how best it should price its products. The objective is to find the best possible action that the agent should take for every product during every time step so that it maximizes its rewards.
The Markov Decision Process is defined by the following tuple: (S, A, R, P, γ), where:

- **S (States):** It describes the status of the item, which includes its shelf life, stock level, and rate of demand.

DOI: 10.17148/IJIREEICE.2026.14419

- **A (Actions):** Represents possible pricing decisions, such as applying different discount levels (e.g., 0%, 10%, 20%, 30%).
- **R (Reward):** Numerical expression indicating the result of an action, taking into account sales income and fines for unsold items or spoilage.
- **P (Transition Probability):** The likelihood of transitioning from one state to another when performing an action.
- **γ (Discount Factor):** Determines the importance of future rewards.

B. System Architecture

- **Data Input Module:** Gathers data from the past and the present, such as product information, expiration dates, daily sales, and inventory levels.
- **State Representation Module:** Shows possible pricing choices, like offering different levels of discounts (0%, 10%, 20%, 30%, etc.).
- **Decision Engine (Q-learning Agent):** Selects optimal pricing actions based on learned Q-values
- **Envirnment Simulator:** Simulate show customers buy products and adjusts inventory levels accordingly based on demand.
- **Feedback Module:** Calculates rewards and updates the Q-table accordingly.

C. State Representation

- Each product is described using a set of key features, including the number of days left before it expires, the current stock available, and its demand level, which can be categorized as low, medium, or high.
- These attributes are converted into a limited set of defined categories to make the learning process more manageable. For instance, remaining shelf life may be grouped into stages such as fresh, mid-stage, or near expiry. This organized format helps the model apply what it learns to similar situations more effectively.

D. Action Space

The set of possible actions includes predefined pricing choices, where each option represents a specific discount applied to the product. For instance:

- No discount (full price)
- Low discount (10%)
- Medium discount (20%)
- High discount (30% or more)

These actions allow the agent to explore different pricing strategies and identify the most effective approach under varying conditions.

E. Reward Function Design

The reward function plays a key role in guiding how the model learns, as it defines what outcomes are considered good or bad. It is structured to achieve two primary goals:

- Increase revenue by assigning higher rewards when products are successfully sold.
- Reduce waste by introducing penalties for items that remain unsold and eventually expire.

In simple terms, the reward can be calculated as the difference between the income generated from sales and the loss caused by unsold products.This approach encourages the model to find a balance between selling items before they expire and maintaining overall profitability.

F.Q-Learning Algorithm

The model uses the Q-learning algorithm to determine the best pricing strategy over time. This approach does not rely on a predefined model; instead, it learns directly from interactions with the environment. Each Q-value reflects the expected return of choosing a particular action in a specific state.

The learning process is based on an update rule:

$$Q(s, a) \leftarrow Q(s, a) + \alpha [r + \gamma \max Q(s', a') - Q(s, a)]$$

where:

- α (learning rate)determines how quickly new information replaces old estimates
- r represents the immediate reward received after taking an action
- s′ is the next state reached after the action
- γ (discount factor) controls how much importance is given to future rewards

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 4, April 2026

DOI: 10.17148/IJIREEICE.2026.14419

To make decisions, the agent follows an ε-greedy strategy, which balances trying new actions and using known good ones:

- With probability ε, it selects a random action to explore
- With probability (1−ε), it chooses the action with the highest estimated value

This method helps the model gradually improve its decisions by learning from both past outcomes and new experiences.

G. Training Process

The model undergoes training by employing either generated or real-life data from supermarkets. The training procedure takes place through several iterations, whereby an iteration is defined as a product's life cycle from stocking up to the expiration of the goods. This includes the following stages during one time step

1. The agent observes the current state
2. Selects a pricing action
3. Receives a reward based on sales outcome
4. Updates the Q-table

Over time, the agent learns optimal pricing strategies for different scenarios.

H. Demand Simulation

The reason for the use of demand simulation model is that the actual customer behavior cannot be predicted. The factors affecting the demand include:Price (higher discounts increase demand)

- Price (the higher the discount offered, the greater the demand)
- Shelf life left
- Variations

This simulation helps create a realistic training environment for the agent.

I. Evaluation Metrics

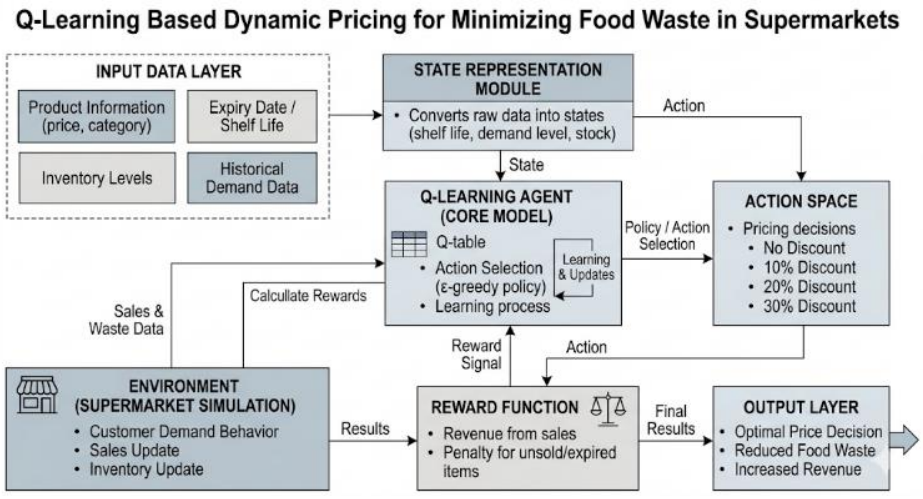
Model Evaluation Criteria:

- **Total Revenue:** The metric used to measure profitability
- **Waste Reduction Rate:** Percentage decrease in the quantity of waste products
- **Inventory Turnover:** The metric evaluates efficiency of stock movement
- **Comparison with Baselines:** Static pricing and rule-based discounting

J. Implementation Tools

Python was used in implementing the proposed model owing to its versatility and extensive applications in machine learning. NumPy and Pandas libraries were utilized in processing data and handling inventory and demand datasets. Q-learning was performed through a Q-table-based method. A simulated environment was set up to mimic demand and product dynamics. Visualization was performed using matplotlib to visualize outcomes such as revenue and waste minimization.

IV. DIAGRAM



IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 4, April 2026

DOI: 10.17148/IJIREEICE.2026.14419

V. RESULTS AND ANALYSIS

Experimental Design The Q-learning-based dynamic pricing scheme was tested in a simulation study by creating a virtual setting for the supermarket. The experiments were performed through various episodes, with each episode being a lifecycle of a product, from its procurement to the expiry period. Some important factors that were involved in the testing process include the value of the learning rate $\alpha = 0.1$, discount rate $\gamma = 0.9$, and an ε-greedy exploration approach with a gradually decreasing ε. The available actions were the discounts applied, including no discount, 10%, 20%, and 30% discounts.

A. Comparative Analysis

The effectiveness of the Q-learning algorithm was benchmarked against two existing solutions:

- **Static Pricing:** A fixed price for all products
- **Rule-Based Pricing:** Pr Predetermined discount schedule based on the product expiration date It is obvious that the Q-learning solution proved to be better compared to the other algorithms applied. The static pricing resulted in huge waste amounts due to its inability to adjust, while the rule-based pricing resulted in more sales but did not optimize profits. The new algorithm changed prices based on market conditions, thus creating a balance in profitability and waste reduction.

B. Revenue Analysis

The Q-learning algorithm resulted in a significant increase in revenue compared to the baseline algorithms. The revenue was increased by using the discount at the most opportune moment, thus avoiding discounting at inappropriate times. The system was observed to have a steady increase in revenue during the various training episodes, indicating that the system was learning from the environment.The Q-learning algorithm resulted in a significant increase in revenue compared to the baseline algorithms. The revenue was increased by using the discount at the most opportune moment, thus avoiding discounting at inappropriate times. The system was observed to have a steady increase in revenue during the various training episodes, indicating that the system was learning from the environment.

C.Waste Reduction Analysis

It is observed that there is a significant reduction in food waste using the proposed approach. When the products were close to their expiration dates, the necessary discounts were applied to encourage their early purchase, thus preventing them from going to waste. The proposed approach is more effective in reducing food waste compared to static pricing, thus proving that it is more practical in a real-world scenario.

D.Advantages of Proposed Method

- A flexible, data-driven pricing approach that adjusts based on changing conditions
- Minimizes the need for manual decision-making by automating pricing adjustments
- Helps maintain a balance between profit generation and reducing waste for better sustainability
- Can be easily applied across a wide range of product categories and retail scenarios

E. Learning Behaviour

The reward graph above indicates that the Q-learning system was able to learn the best pricing strategies. At the initial state, the rewards were not consistent due to the exploration of the strategies by the system. However, the results were consistent and improved with time, showing that the system converged to a point where the results were optimal.

F. Summary of Results

Method	Revenue Level	Waste (%)
Static Pricing	Low	High
Rule-Based	Medium	Medium
Q-Learning	High	Low

G. Key Observations

- Q-learning effectively adjusts to shifts in customer demand over time
- Dynamic pricing helps improves both revenue generation and waste reduction
- The model strikes a balance between maximizing profit and sustainability goals
- Reinforcement learning offers a scalable and practical solution for real-world retail systems

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 4, April 2026

DOI: 10.17148/IJIREEICE.2026.14419

VI. FUTURE WORK

The suggested dynamic pricing model based on Q-learning shows promise in cutting down on food waste and boosting revenue. However, more research is needed to make it more effective and useful in the real world. One significant avenue is the expansion of the model to Deep Reinforcement Learning, including Deep Q-Networks (DQN), which can manage larger and more intricate state spaces without requiring discretization. This would let the system pick up on more specific details, such as changes in demand over time and how customers behave in real time.Integrating real-time data from supermarket systems is another useful way to make things better in the future. Using real-time data like point-of-sale transactions, stock level updates, and customer foot traffic would help the pricing strategy respond more quickly and accurately to what is really going on in the store. Also, Also,using the model with more advanced methods for predicting demand could make it work even better by helping to predict how customers will buy things and overall trends in demand.at the moment, the model is being tested in a simplified simulation environment. Future research should focus on putting it to the test in real retail settings. This would give a better idea of how well it works in practice and help find real-world problems like scalability, integration with current systems, and getting stakeholders to accept it.

Future work could also look into making the action space bigger so that prices can change more freely and continuously instead of being stuck at fixed discount steps. Another helpful change would be to offer personalized pricing strategies that take into account what customers like and what they have bought in the past. This could help boost overall sales.

Also, using new technologies like IoT sensors and smart inventory management systems could give you more accurate and up-to-date information about the state of your products. This would help people make better and faster decisions about prices. All of these improvements would make the system stronger, more flexible, and better for real-world smart retail uses.

VII. DISCUSSIONS

The results of the proposed Q-learning–based dynamic pricing model indicate a clear improvement over traditional static and rule-based pricing strategies. One of the key reasons for this performance gain is the model's ability to adapt pricing decisions based on real-time factors such as remaining shelf life, demand levels, and inventory status. Unlike static pricing, which treats all products equally regardless of their condition, the proposed approach adjusts prices dynamically, increasing the likelihood of selling items before expiration.

The model shows a strong ability to handle the trade-off between revenue maximization and waste reduction. In the initial stages of the product life cycle, it maintains a relatively high price to ensure profit maximization when the demand is relatively stable. When the product is near expiration, it increases the rate of discounts to ensure a faster rate of sales and minimize waste. This dynamic behavior ensures the sale of the product at optimal price points rather than pre-defined discount rules.

The flexibility and scalability of the suggested model can be seen when compared to the rule-based approach. In the rule-based approach, rules are set according to the situations. However, the rules may not generalize well for different products. In the Q-learning model, the policy is constantly updated according to the outcome, which makes the model robust.

Despite these benefits, there are a few limitations to this model. For instance, a simulated environment may not accurately depict real-world customer responses. Moreover, the demand responses considered are relatively simple. Further, the discretized action space allows for only a specific discount amount for price control, which may not necessarily lead to optimal results. This shows that although this model works well, there are still a few improvements that need to be made for real-world implementation. Overall, it can be seen that the discussion shows that the proposed approach works well to remove the limitations of traditional approaches to pricing strategies. The approach introduces adaptability and learning capabilities. Further, it shows that reinforcement learning can improve efficiency as well as sustainability in retail systems.

VIII. CONCLUSION

This study introduced a Q-learning-based dynamic pricing model to tackle the problem of food waste in supermarkets. The pricing process was modeled as a Markov Decision Process, allowing the system to learn effective pricing strategies based on factors such as product shelf life and demand. Unlike traditional fixed pricing methods, this approach adjusts prices dynamically, leading to better and more timely decisions.

DOI: 10.17148/IJIREEICE.2026.14419

The results show that the model can reduce food waste while maintaining or even improving overall profitability. By applying suitable discounts at the right time, it helps ensure that perishable items are sold before they expire instead of being wasted.

Overall, this work demonstrates how reinforcement learning can be applied to real-world retail challenges. It also provides a strong base for developing smarter and more automated pricing systems. Future work can focus on using real-time data and implementing the model in real supermarket environments.

REFERENCES

[1]. K. Talluri and G. van Ryzin, *The Theory and Practice of Revenue Management*. New York, USA: Springer, 2004.
[2]. R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.
[3]. C. J. C. H. Watkins and P. Dayan, "Q-learning," *Machine Learning*, vol. 8, no. 3–4, pp. 279–292, 1992.
[4]. S. Agrawal, V. Avadhanula, V. Goyal, and A. Zeevi, "Thompson Sampling for Dynamic Pricing," *Operations Research*, vol. 67, no. 6, pp. 1503–1525, 2019.
[5]. M. Ferreira, R. M. de Souza, and L. A. Barroso, "Dynamic Pricing for Perishable Products Using Machine Learning," *Journal of Retailing and Consumer Services*, vol. 47, pp. 1–10, 2019.
[6]. FAO, "Global Food Losses and Food Waste – Extent, Causes and Prevention," Food and Agriculture Organization of the United Nations, 2011.
[7]. J. F. Shapiro, *Modeling the Supply Chain*, 2nd ed. Boston, MA, USA: Cengage Learning, 2007.
[8]. H. Chen and S. Simchi-Levi, "Pricing and Inventory Management," in *Handbooks in Operations Research and Management Science*, vol. 12, Elsevier, 2004, pp. 275–300.
[9]. Y. Ye and D. Zhang, "Dynamic Pricing and Inventory Control for Perishable Products," *European Journal of Operational Research*, vol. 198, no. 3, pp. 858–868, 2009.
[10]. OpenAI Gym, "A Toolkit for Developing and Comparing Reinforcement Learning Algorithms," [Online].