Production Planning for Pineapple Canned with Reinforcement Learning

Payungsak Klasantia[1]* Noppadol Amm-Dee[1]* and Chidchanok Choksuchat[2]

[1]Department of Industrial Technology Management, Faculty of Industrial Technology,
Muban Chombueng Rajabhat University, Ratchaburi, 70150

[2] Division of Computational Science, Faculty of Science, Prince of Songkla University, Songkhla, 90110

*__Corresponding author__: payungsakklasantia@gmail.com, noppadolamd@mcru.ac.th

Received: 22 September 2024/ Revised: 21 February 2025/ Accepted: 24 February 2025

Abstract

Pineapple is a significant economic crop in Thailand, and the pineapple processing industry is crucial for farmers, manufacturers, and customers. Production planning is challenging due to the complexity of customer demands and the increasing uncertainty of fresh pineapple yields. The researcher studied and experimented with algorithms for production planning for canned pineapple using reinforcement learning. The objective was to optimize production planning by finding the best values for just-in-time scheduling through reinforcement learning, based on the Markov Decision Process (MDP) used for sequential decision-making. After analyzing the problem and recognizing that production planning has such characteristics, the researcher proceeded as follows: 1) Designed a dataset from case study data and defined the objective function. 2) Developed a reinforcement learning model using the Advanced Actor Critic (A2C) algorithm to create the production plan for the case study. 3) Tuned the model's parameters, trained the model, and tested it. 4) Evaluated the model and found that the reward or the defined objective function increased by at least 40% compared to the initial model. Additionally, the sellable products' readiness improved by 120%, and the discrepancy between expected and actual returns decreased by 19% compared to the initial machine learning model.

Keywords: Production scheduling, reinforcement learning, Markov Decision Process, production planning, machine learning

Introduction

Pineapple is one of Thailand's important economic crops because Thailand can generate income from exporting pineapples, processed pineapple products, and canned pineapples. The Food and Agriculture Organization of the United Nations [1] has estimated the value of processed pineapple products of Thailand as the world's third largest pineapple exporter in 2021, as shown in Figure 1.

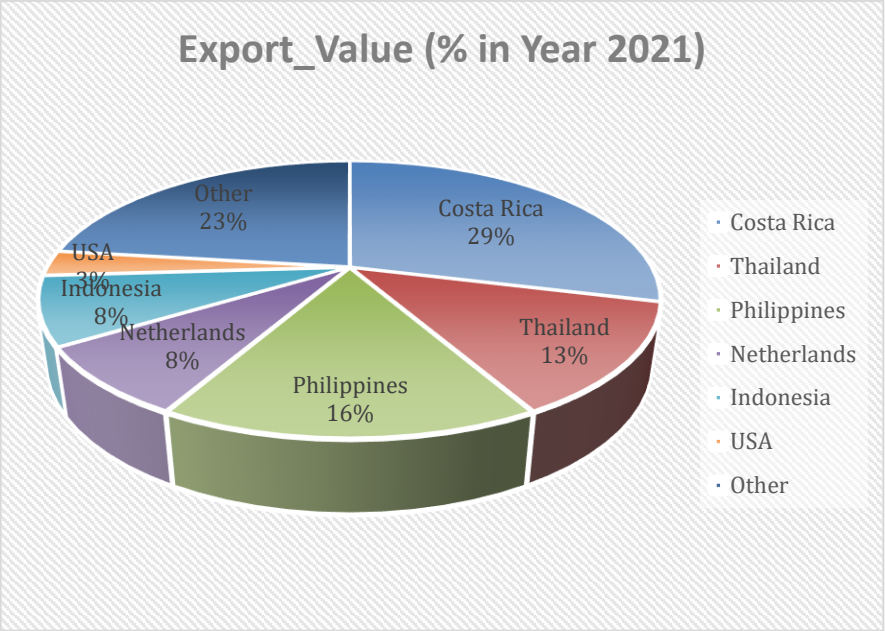

Figure 1 The percentage of pineapple product export value in 2021.

The pineapple processing industry can generate an average of 10 billion baht per year for the country from exports, considering data from each year from 2018 to 2022, referring to the database from the Office of Agricultural Economics, as shown in Table 1.

Table 1 Value and volume of exports of pineapples packed in airtight containers, 2018-2022

Year	Value (in million baht)	Value (in tonnes)
2018	10,418	313,142
2019	8,402	250,477
2020	7,429	150,755
2021	10,124	201,480
2022	12,144	211,321

Research findings [2, 3] indicate that the pineapple industry involves a link between farmers, processing plants, and customers. However, pineapple processing facilities face several challenges, including production inconsistencies that hinder accurate production planning and result in high error rates. Additionally, the cost of Thai pineapple products remains higher than that of competitors in the global market. In response to these issues and in alignment with the Pineapple Strategy 2017-2026, strategic policies have been established for processing plants to enhance production efficiency and decrease costs. Effective decision-making in production planning is crucial for helping factories meet production targets, maintain appropriate inventory levels, and prioritize products to satisfy fluctuating customer demand, ultimately minimizing total costs. Researchers have proposed various tools to address production planning challenges, such as enhancing sustainable supply chain efficiency for banana production using a mixed-integer linear

programming model [4], and employing a similar model for optimizing production schedules in canned fish factories, achieving near-optimal results in real-world scenarios [5]. Moreover, ongoing studies focus on developing artificial intelligence systems to autonomously optimize inventory management, resulting in lower overall costs compared to traditional systems. Reinforcement learning is a branch of machine learning focused on training agents to make decisions that maximize long-term rewards. These agents interact directly with their environment, using trial-and-error methods to discover optimal strategies based on feedback from their actions, including rewards and penalties. Previous studies have demonstrated that reinforcement learning has been effectively utilized to address production planning uncertainties across various industries [6-12]. This approach is rooted in the Markov Decision Process (MDP) framework proposed by Bellman in 1957 [13]. For instance, in 2020 [14], research applied a Deep Neural Network algorithm in Reinforcement Learning concept to tackle supply chain challenges in the Beer Game model developed by MIT, which encompasses the interactions between manufacturers, distributors, sellers, and retailers. The focus of this research was on minimizing supply chain costs and enhancing decision-making in real-world scenarios. By employing the Deep Q Network algorithm, specifically the Sharpened Reward DQN (SRDQN), the study was able to achieve strong results using a smaller data set, outperforming traditional base stock policies. Additionally, research addressing production system issues in complex manufacturing environments has highlighted the significant impact of order dispatching and maintenance management on overall production efficiency [14-18], especially when real-time decision-making is essential. Scholars have advocated for the use of reinforcement learning in this context due to its capability for autonomous operation, environmental adaptation, and optimal value discovery. Through reward and penalty analysis, one study constructed a model using the Trust Region Policy Optimization algorithm to evaluate reinforcement learning outcomes. The findings revealed that work dispatching to the production line using reinforcement learning achieved machine utilization rates exceeding 90%, significantly surpassing the efficiency of the First In First Out (FIFO) heuristic method. Furthermore, the overall production process time was reduced in comparison to FIFO. This indicates the potential of reinforcement learning to effectively resolve various industrial challenges. Researcher [19] studied the RL model using Q-Learning combined with the Knowledge-Driven Greedy Algorithm (KDQRL) for thermo-mechanical finite element analysis (FEA), comparing it with other methods such as Modified Low-Cost Search (MLCS), GA, and Exhaustive Search, which resulted in a 71% reduction.

This research aims to identify the most suitable values for just-in-time production planning in the canned pineapple processing industry, utilizing reinforcement learning to assess its effectiveness in various scenarios. Given the importance of this topic, the study seeks to provide insights into resolving production scheduling challenges commonly faced in the conventional pineapple processing industry, which often relies on the experience of skilled labor. By leveraging machine learning, specifically reinforcement learning, the research will extract insights from high-level data, such as forecasting and actual customer orders, to determine the optimal actions based on maximizing returns across diverse situations.

Materials and methods

1. Dataset Design and Target Function Definition

Reinforcement Learning uses feedback from interaction with the environment to learn the policy for the Markov Decision Process (MDP) through trial and error. The model requires extensive training, initially making mistakes, and learning which actions lead to optimal behavior. This approach is also applicable to more complex decision-making problems. The RL model uses the A2C algorithm, consisting of two components:

Actor: Learns a stochastic policy and selects actions with the highest probability of maximizing future rewards. Critic: Approximates the value of the current policy by estimating the expected sum of future discounted rewards. The system state, containing decision-making information such as inventory levels, actual demand, forecast data, the current schedule, and time, is passed from the environment to the algorithm. The agent uses a Deep Neural Network (DNN) to map states to actions, represented as a stochastic policy that provides a probability distribution of possible actions for each state. The reward function evaluates actions taken at specific states and times, aligning with the objective function to optimize the planning horizon. A diagram illustrates the A2C-based reinforcement Learning model, as shown in Figure 2.

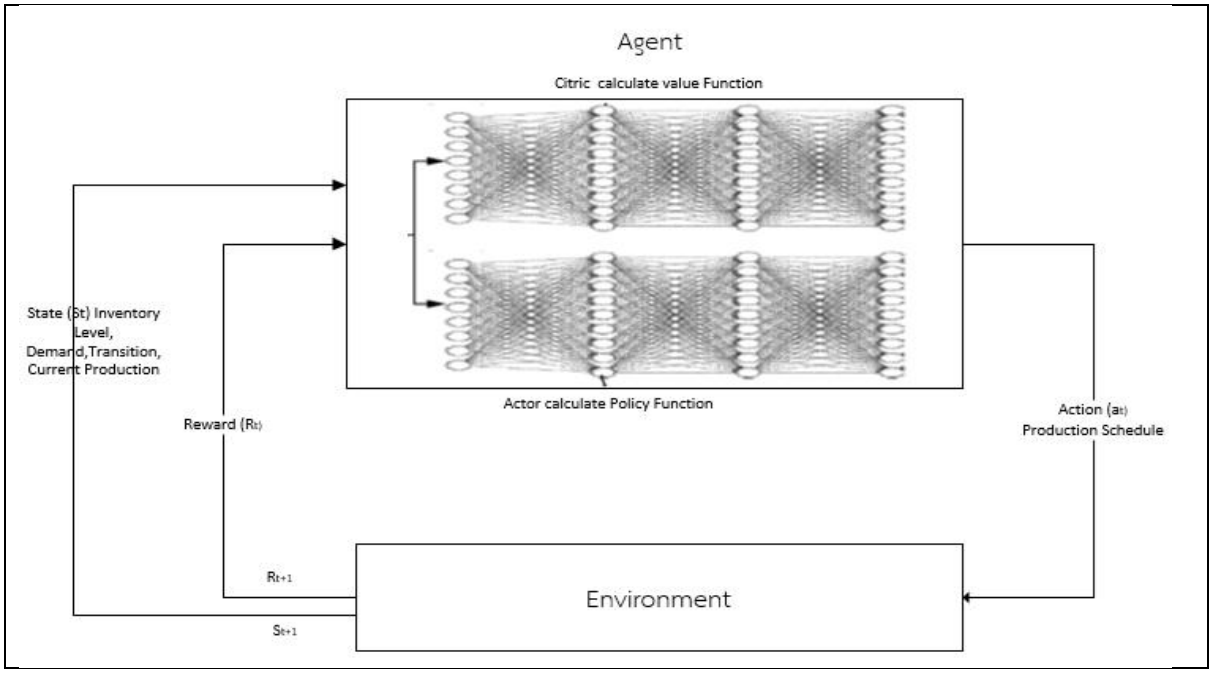

Figure 2 The diagram of Reinforcement Learning model with A2C algorithm for Canned Pineapple production planning

From the study of the general conditions and problems of the case study, the researcher can design a dataset related to the production planning and scheduling system. An important dataset example includes finished goods data (Table 2), inventory data (Table 3), transition probability data (Table 4) is a probability distribution over next possible successor states, given the current state, and customer order data (Table 5). The customer order data consists of the following attributes: order number, delivery date, ordered products, quantity ordered, profit per unit, creation date, and order type, among others.

Table 2 Finished Good Data

FGCode	Description	UOM	Safety Quantity	Sale Price	Cost
A	FANCY SLICE 2T (F200)CASE 0.5	EA	10000	614	584
B	FANCY CHUNKS (F200)CASE 0.5	EA	10000	620	600
C	TIDBITS/CUBES (F200)CASE 0.5	EA	10000	638	606
D	CRUSH (F200)CASE 0.5	EA	4000	610	590

Table 3 Inventory

CalendarDay	FG_Code	Inventory_Balance
1	A	100000
1	B	20000
1	C	150755
1	D	201480

Table 4 Transition probability Data

From/To	To_A	To_B	To_C	To_D
A	0	100	100	200
B	0	0	100	200
C	0	0	0	300
D	0	0	0	0

Table 5 Value and volume of pineapples packed for demand from customer

Doc_num	Doc_create	Planned_gi	FG_code	Order_qty	Var_std	Cust_segment
9791	122	127	D	3024	31	1
9793	108	113	B	5040	30	1
9803	120	127	A	6300	34	1

Definition of variable of the model

Set and index variables

i = product i when $i = 1, 2, 3, ...I$

j = product j when $j = 1, 2, 3, ...J$

m = machine m when $m = 1, 2, 3, ...M$

t = index for time period t when $t = 1, 2, 3, ...T$

n = index for order number from customer n when $n = 1, 2, 3, ...N$

Continuous variables

M_{it} = quantity of product i produced at time t when $i = 1, 2, 3, ...I$

when $t = 1, 2, 3, ...T$

F_{it} = finished inventory level of product i at time t

when $i = 1, 2, 3, ...I$

when $t = 1, 2, 3, ...T$

O_{itn} =Sales from sales of product i for orders n made in the initial period until time $t = t$

when $i = 1, 2, 3, ...I$

when $n = 1, 2, 3, ...N$

when $t = 1, 2, 3, ...T$

Binary variables

x_{imtn} =Binary variable equal to 1 means that product i is shipped according to order n at machine m at time t, otherwise 0

y_{imt} =Product i is ordered at machine m at initial time t, when y_{imt} equals 1 means that product is ordered, otherwise 0

Z_{ijt} = When Z_{ijt} = 1 means that a change in the state of the product has occurred, otherwise 0

Parameter

D_{itn} = Demand for product i at time t for order from customer n

t_n = Due date t of order from customer n

G_{in} = Standard profit minus delay loss on each order day

δ_{it} = Product change from production plan from product i to product j

$C_{im}Max$ = Maximum production capacity of product i

Objective Function or Reward Function to be used in evaluating the representative of the model

$$Z = \sum_{n}\sum_{i}\sum_{t} G_{in}O_{itn} - \eta\sum_{i}\beta F_{it} \quad (3)$$

Z is the competitiveness level of the business, it is the cumulative profit from sales of goods minus the coefficient of return on goods multiplied by the average cost.

G_{in} is the profit from sales of product i for order n.

O_{int} is the sales from sales of product i for order n in the period from the beginning to time $t = T$.

η is the coefficient of return on goods, calculated by dividing profit by the cost of goods, giving a value of 0.05.

F_{it} is the inventory level of product i at time t. The constraint on the level of available inventory for sale.

$$F_{it} = F_{it-1} + M_{it-2} - \sum_{j}\delta_{ij}Z_{ijt-2} - \sum_{n}\sum_{t\geq tn} O_{itn} \forall i,t \quad (4)$$

Equation 4 shows that the inventory level of product i at time t is the product inventory level from the previous period $t-1$ plus the volume produced two days ago, M_{it-2}, minus the change in product from the planned production at time $t-2$ $\delta_{it}Z_{ijt-2}$, and minus the volume of product i sold to customer n at time t (O_{itn}).

Sales constraints

$$O_{itn} = D_{itnn}U_{itn}\forall n,i,t \geq t_n \quad (5)$$

$$\sum_i \sum_{t \geq t_n} x_{itn} \leq 1 \forall n \quad (6)$$

Equations 5 and 6 represent the sales volume that can satisfy the demand for a product at a given time.

D_{itn} represents the demand for product i for order n at a given time t_n

x_{itn} is a binary variable equal to 1, meaning that product i is delivered according to order n at time t, and 0 means that it cannot be delivered at the given time.

$t \geq t_n$ means that t must be greater than or equal to t_n, so the quantity of product that can fulfill the customer order.

Production limitations

$$\sum_i v_{it} = 1 \forall t \quad (7)$$

$$0 \leq M_{it} \leq C_{im}^{Max} \forall i,t \quad (8)$$

Equation 7 shows that if product i is produced at time t, where v_{it} is a binary variable, then if product i is produced, v_{it} equals 1, but if it is not produced, then v_{it} equal 0. Each production line can produce the product and the production capacity is equal to $C_{im}^{\max}$ for the quantity of production of that product i at machine m not to be greater than the production capacity of the production line, as shown in Equation 8.

2. Developing the model

The development of the production scheduling management model involves utilizing reinforcement learning through deep neural networks, specifically employing the Advanced Actor-Critic (A2C) algorithm. The researcher fine-tuned the parameters of the production scheduling model, as well as the training and testing models, using an enhanced development program from PVC Development, further supported by Python Version 3.11.1 on a Dell Latitude 3340 with an Intel Core(TM) i5-4200U processor, 4 GB RAM, and Windows 10 Pro.

The model was created specifically for the canned pineapple industry, using data from a case study of a company operating in Prachuapkhirikhan Province. This case study involved six products requiring production planning and utilized customer purchase order data from 2022, alongside the production capacity of the Number 2 canned production line. Additionally, the model accounted for the variability of new products based on the quality of fresh pineapples received at the factory and outlined the required planning period.

Results and discussion

The initial hyperparameters for the production planning model were defined, alongside custom hyperparameters, enabling the creation of a production schedule based on these parameters. The researcher generated a production schedule using the initial hyperparameter model, represented graphically

with the products on the vertical axis and production dates on the horizontal axis. Moreover, the results of the total reward from the initial hyperparameter model were plotted against the teaching episodes of the model, as illustrated in Figures 3 and 4.

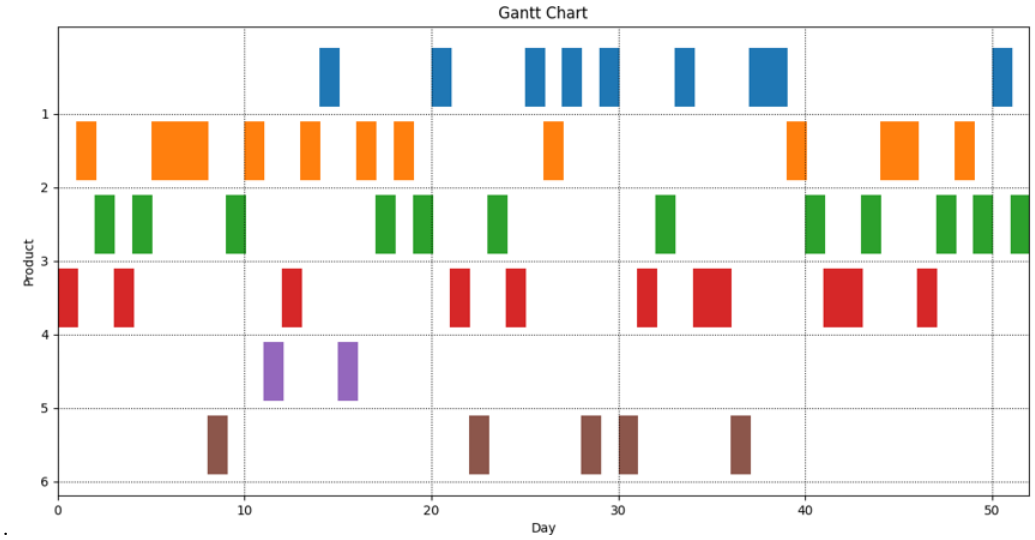

Figure 3 Production schedule from the initial hyperparameter model

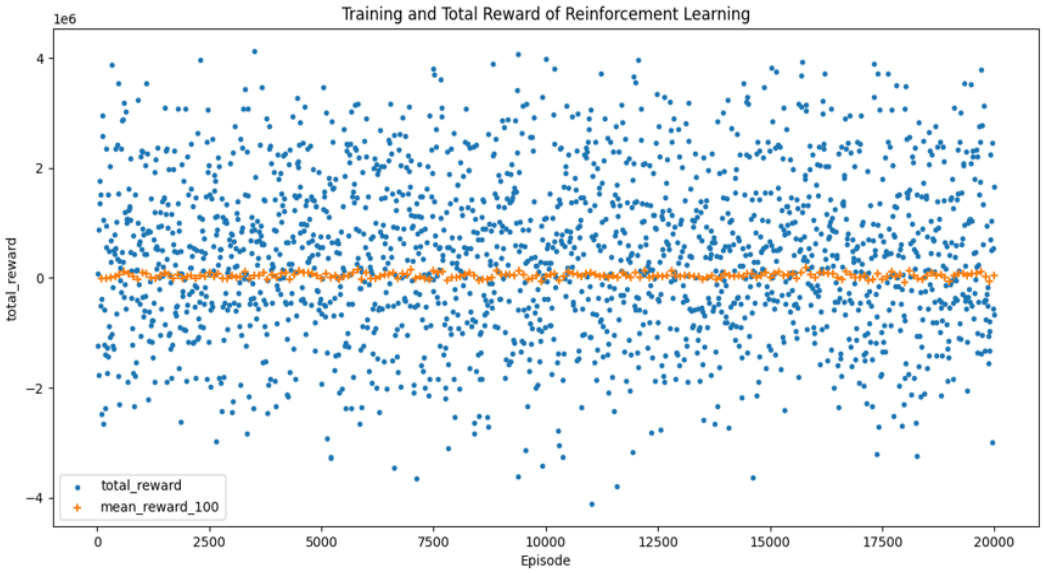

Figure 4 The results of the total return of the initial hyperparameter model with model training rounds.

The researcher adjusted the model by adjusting the hyperparameters and using the same training data as shown in Figure 5 which is the production schedule from the tuned hyperparameter model. Figure 6 is the graph showing the total returns of the tuned hyperparameter model. Figure 7 is the graph showing the returns in other parts including total returns, product cost, penalty effect for late delivery, and returns from on-time delivery. Figure 8 is the graph showing the products ready for delivery.

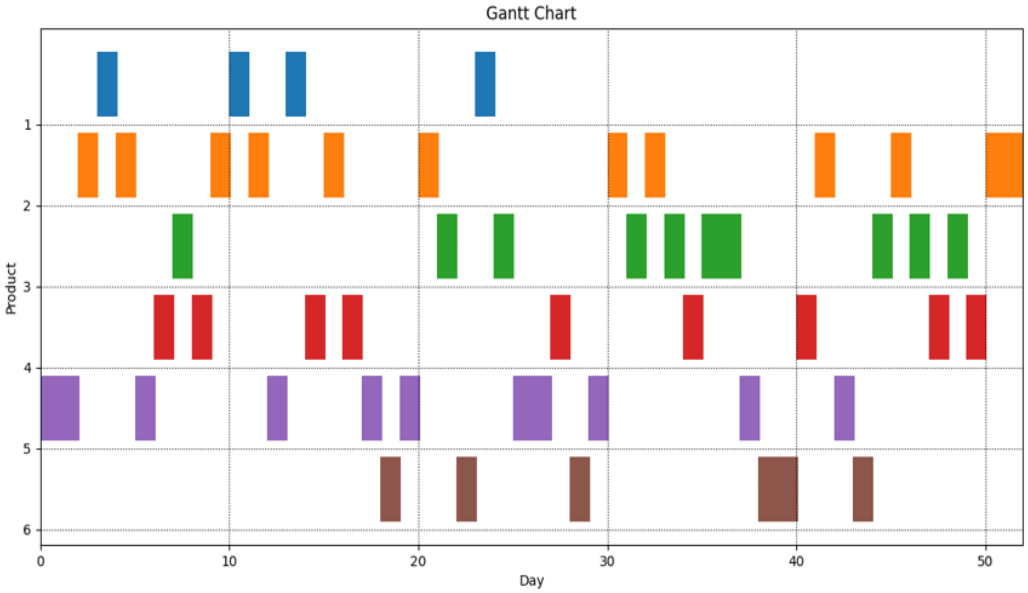

Figure 5 Production schedule from the optimized hyperparameter model.

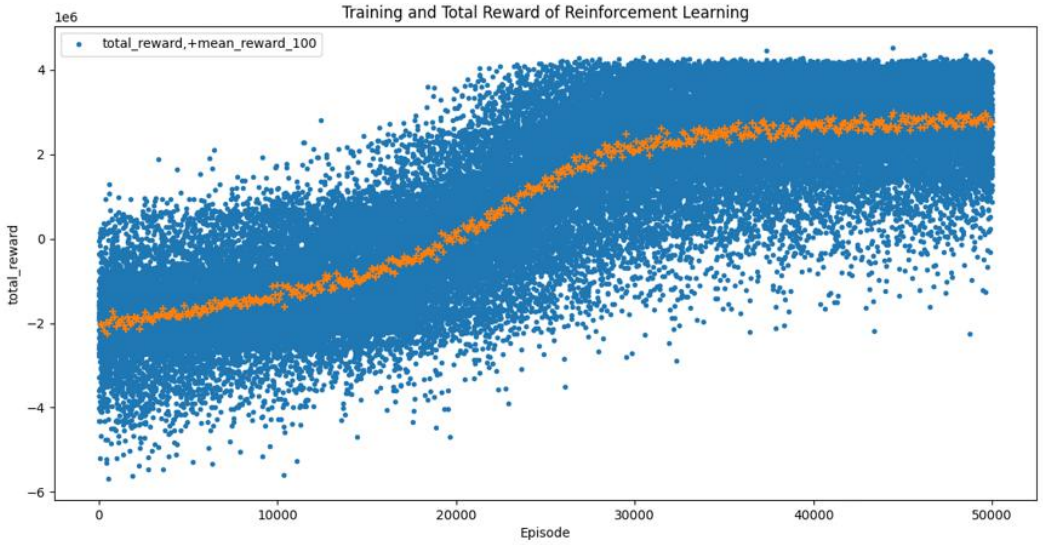

Figure 6 The total returns of the adjusted hyperparameter model

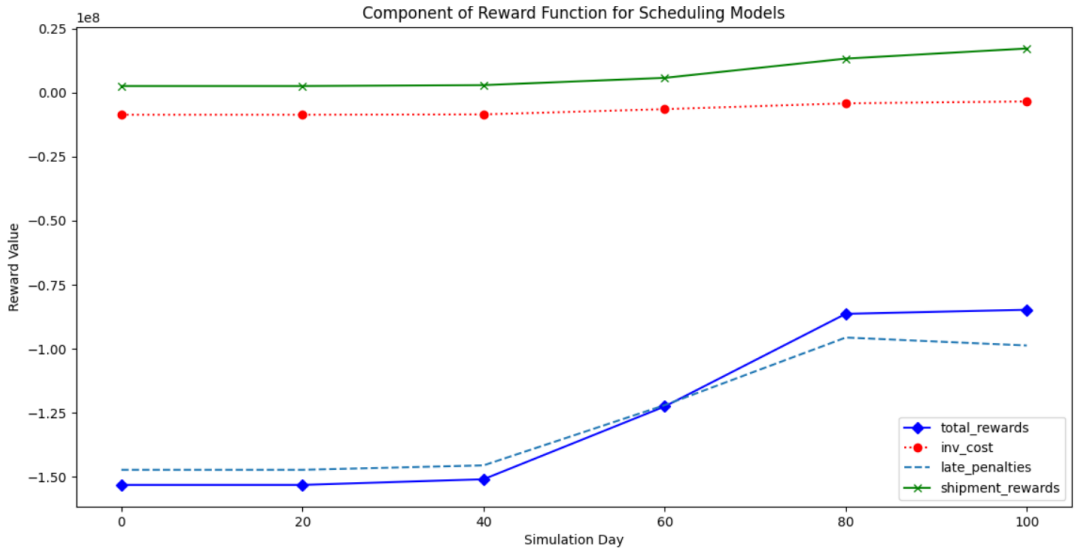

Figure 7 Returns in other sections

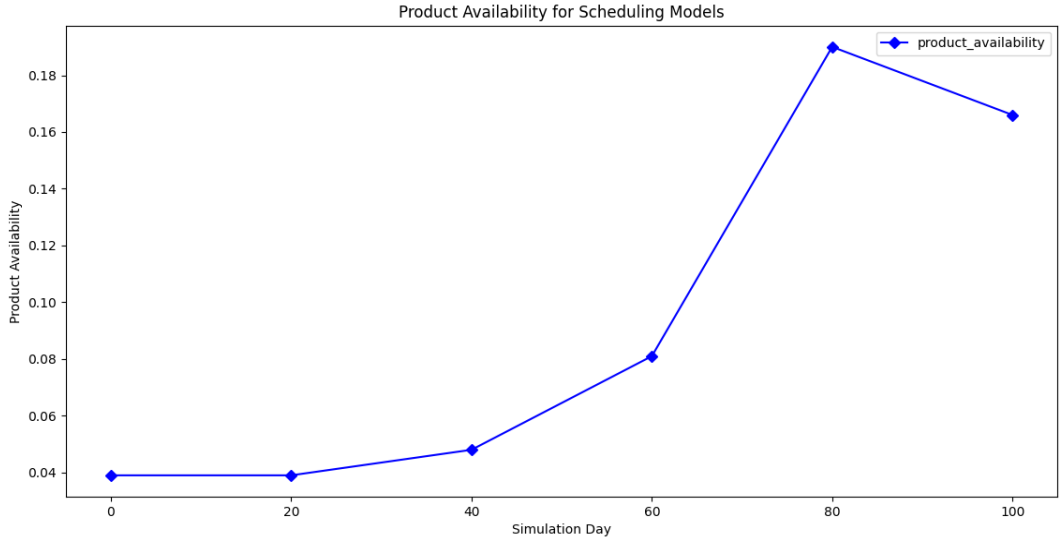

Figure 8 Ready-to-ship products

The research results on addressing the production planning challenges in the canned pineapple industry revealed significant improvements in the models before and after adjustments. The model representatives demonstrated a progressive learning curve, with their performance steadily increasing and stabilizing after approximately 40,000 training iterations. The adjusted model significantly enhanced operational performance, leading to a 44% increase in overall returns, while the error rate in return calculations decreased by 52%, ensuring greater accuracy in financial assessments. Additionally, the return on sales improved by 29%, reflecting higher profitability, and the inventory of ready-to-ship products expanded by a factor of 1.77, enhancing availability and responsiveness to market demand. These findings are summarized in Table 6.

Table 6 Results of Evaluation

Evaluation Items	Pre-Model(A)	Post-Model(B)	Difference C=(B-A)*100/A
Return Value	-153.21	-16.77	44%
Level of Return Deviation	0.11	0.24	52%
Return on Sales to Customer	13.36	17.32	29%
Product Availability	0.04	0.17	325%

Mixed-Integer Linear Programming (MILP) model is one popular mathematical model that the researchers experimented with and compared against Posted training RL model. The results for Total Reward, Inventory Cost, Late Penalty, and Shipment Reward are shown, Table 7 demonstrates the comparison between RL and MILP. The table reveals that the RL model outperforms the MILP model by about 50.2%, with RL's inventory cost being approximately 55% lower than MILP's, and RL's Product Availability improving by 50.1%.

Table 7 Results of Evaluation comparison between Post RL and MILP

Evaluation Items	Post-RL Model(A)	MILP(B)	Percentage Difference C=(A-B)*100/B
Return Value	-84.77	-160.77	47%
Return on Sales to Customer	17.32	16.32	6.7%
Product Availability	0.17	0.13	30%

The parameters used for training the model are as follows: batch size = 1000, discount factor (γ) = 0.99, number of hidden nodes = 512, number of hidden layers = 64, activation function = ELU, learning rate = 0.95, actor learning rate = 5e-07, critic learning rate = 1e-05, bias = True, and beta = 1e-05.

Researcher presented a sensitivity analysis of hyperparameters. Changes in hyperparameters impacted the model's performance. It was found that as the number of iterations increased, the effect on returns became more significant. Additionally, adjusting the number of Layer and Node in Neural Network inspections resulted in higher returns show as in Figure 9

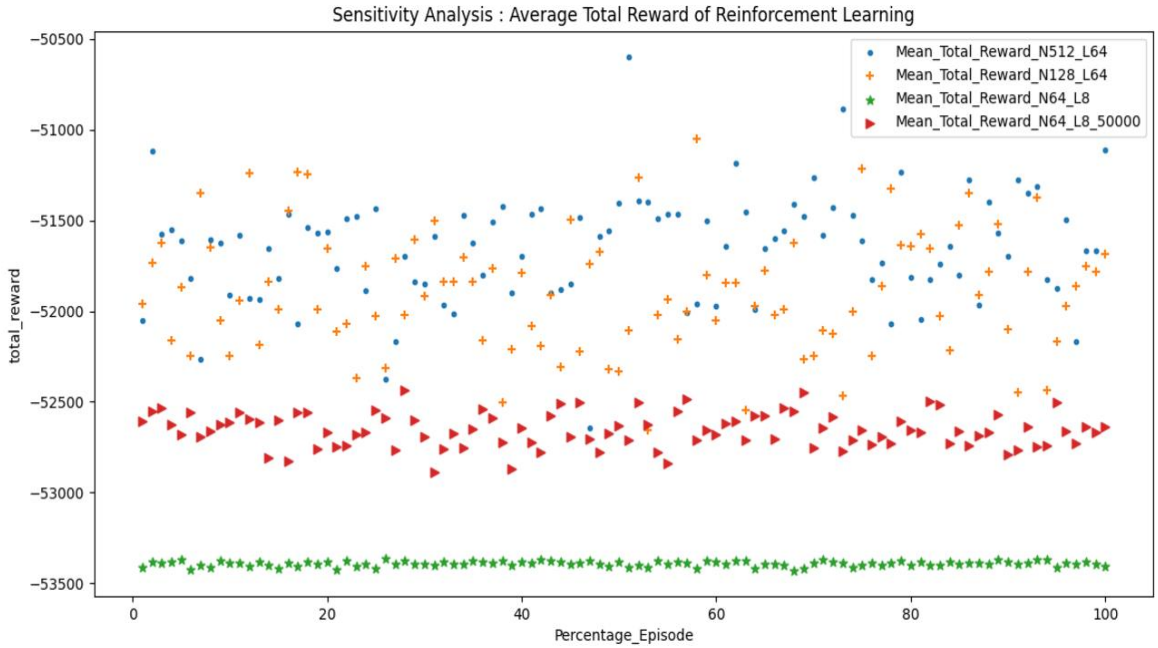

Figure 9 Sensitivity Analysis of RL Model

Conclusion

Production planning plays a crucial role in the manufacturing industry, particularly within the agricultural sector, where uncertainty levels are typically higher. Effective production planning that yields increased returns is regarded as a strong organizational strategy. The analysis of the production plan developed using the adjusted reinforcement learning model demonstrates improved returns compared to previous efforts, along with positive evaluations across other dimensions while adhering to business constraints related to canned pineapple production.

The long-term impact of the reinforcement learning model on the canned pineapple industry is significant. Economically, the model has the potential to stabilize production costs, enhance market responsiveness, and provide consistent financial benefits to stakeholders. [20] Environmentally, its ability to optimize production schedules can reduce resource wastage, energy consumption, and overproduction, aligning with sustainability goals. Socially, improved efficiency may benefit farmers by ensuring stable demand and fair pricing while fostering consumer satisfaction through reliable product availability.

The scalability of the model was also explored, demonstrating its potential for broader applications beyond the canned pineapple industry. The model's flexibility allows it to handle larger problem instances and adapt to more complex production planning scenarios, making it suitable for other agricultural and manufacturing contexts. However, challenges such as computational demands and data management must be addressed to fully realize its scalability.

Future research should focus on integrating the reinforcement learning model with other decision-making processes within the pineapple processing industry, such as inventory control, supply chain management, and quality assurance. This integration could create a unified framework that enhances decision-making efficiency and resilience, ensuring comprehensive operational improvements. Additionally, testing alternative algorithms and incorporating constraints such as raw material availability, labor resources, and storage capacity can further refine the model.

The implementation of reinforcement learning in real-world production environments requires addressing challenges like ensuring high-quality data collection, computational resource availability, and seamless integration with existing systems. Engaging industry stakeholders and demonstrating tangible benefits through pilot programs will be essential for gaining acceptance. Finally, ethical considerations such as potential job displacement and data privacy concerns must be proactively managed to ensure responsible AI deployment in production planning.

References

1. Food and Agriculture Organization of the United Nations. FAOSTAT [Internet]. 2023 [cited 2023 May 6]. Available from: https://www.fao.org/faostat/en/#data/WCAD.
2. Tantipipaphong K. Guidelines for promoting agricultural product processing for export. Thesis, National Defense Course, National Defense College. Thailand; 2020.
3. Saengchan S. Development of Marketing Strategy of Thai Canned Pineapple Exporters. Master of Business Administration Thesis, Burapha University College of Commerce. Thailand; 2017.
4. Chantaros P, Klaychey R, Limpianchob C. Optimization of a Sustainable Supply Chain Planning System for Cultivated Banana Production with a Mixed-integer Linear Programming Approach. TJOR 2023;11(1):21-3.
5. Georgiadis GP, Mario Pampin B, Cabo DA, Georgiadis MC. Optimal production scheduling of food process industries. Comput Chem Eng 2020;134:106682.
6. Hubbs CD, Li C, Sahinidis NV, Grossmann IE, Wassick JM. A deep reinforcement learning approach for chemical production scheduling. Comput Chem Eng 2020;141:106982.

7. Kumar A, Dimitrakopoulos R, Maulen M. Adaptive self-learning mechanisms for updating short-term production decisions in an industrial mining complex. J Intell Manuf 2020;31(6):1795–1811.
8. Guo F, Li Y, Liu A, Liu Z. A reinforcement learning method to scheduling problem of steel production process. J Phys Conf Ser 2020;1486:072035.
9. Woo JH, Kim B, Ju S, Cho YI. Automation of load balancing for Gantt planning using reinforcement learning. Eng Appl Artif Intell. 2021;101:104226.
10. Kemmer L, Kleist H, Rochebouet D, Tziortziotis N, Read J. Reinforcement learning for supply chain optimization. EWRL 2018;14(10):1-9.
11. Schwung D, Schwung A, Ding SX. Actor-critic Reinforcement Learing for Energy Optimization in Hybrid Production Environment. IJC 2019; 18(4):360-71.
12. Geevers K. Deep Reinforcement Learning in Inventory Management. Master Thesis, Industrial Engineering and Management, University of Twente. Netherlands; 2020.
13. Bellman R. A Markovian decision process. J Math Mech. 1957;6(5):679-84.
14. Oroojlooyjadid A, Nazari M, Snyder L, Takáč M. A Deep Q-Network for the Beer Game: A Deep Reinforcement Learning algorithm to Solve Inventory Optimization Problems [Internet]. 2020 [cited 2024 Nov 01]. Available from: https://arxiv.org/abs/1708.05924.
15. Kuhnle A, Kaiser JP, Theiß F, Stricker N, Lanza G. Designing an adaptive production control system using reinforcement learning. J Intell Manuf. 2021 ;32(3):855-76.
16. Martínez JY, Coto Palacio J, Nowé A. Multi-agent reinforcement learning tool for job shop scheduling problems. Commun Comput Inf Sci 2020;1173:3-12.
17. Lang S, Behrendt F, Lanzerath N, Reggelin T, Müller M. Integration of deep reinforcement learning and discrete-event simulation for real-time scheduling of a flexible job shop production. Proc 2020 Winter Simul Conf (WSC) 2020;3057-68.
18. Chang J, Yu D, Hu Y, He W, Yu H. Deep reinforcement learning for dynamic flexible job shop scheduling with random job arrival. Processes 2022;10(4):760.
19. Romero-Hdz J, Saha BN, Tsutsumi S, Fincato R. Incorporating domain knowledge into reinforcement learning to expedite welding sequence optimization. Eng Appl Artif Intell 2020;91:103612.
20. Modrak V, Sudhakarapandian R, Balamurugan A, Soltysova Z. A review on reinforcement learning in production scheduling: an inferential perspective. Algorithms 2024;17(8):343.