SDP Research Report - Template

Project Name	Reinforcement Learning on Sales & Distributions
Faculty Supervisor Name	Mr. Samir Mammadov
Team Lead Name & ID	Laman Panakhova 16882 Introduction & Review
Team member #1 Name & ID • What sections did this person contribute to?	Nargiz Aghayeva 16042 Proposed Solution & Conclusion
Team member #2 Name & ID • What sections did this person contribute to?	Aysu Azammadova 16113 Expected Outcomes & Conclusion
Team member #3 Name & ID • What sections did this person contribute to?	Leyla Eynullazada 18033 Ethical Considerations & Review

Word count guide: 2000-2500

Table of contents

Over the last ten years, Supply Chain Management (SCM) grew from an operations proce ss to a smart, data-based system that is essential for world efficiency and sustainability. New SC M leverages analytics and technology to improve every step of it, dealing with fluctuating marke ts, unpredictable demand, and sustainability goals. Standard optimization methods, however, hav e trouble dealing with uncertainty and complex interdependencies. Reinforcement Learning (R L), which is an area of AI, offers a solution through learning actions optimally through interactio n and feedback. In contrast to static models, RL learns in real time and can be ideally applied to dynamic SCM problems such as pricing, inventory control, and resource allocation for rapidly ch anging environments.

However, numerous limitations still persist. A large part of the work hinges on abstracted or simulated conditions that are not able to recreate real-world complexity and uncertainty of real -world supply chains. Moreover, there is a lack of theoretical advancements being translated into industrial practice due to data shortages, scalability, and computationally costly nature of RL lear ning. The most important research question guiding this study is: "How can reinforcement learni ng best be applied to decision-making in high-uncertainty, complex supply chain situations with i nterpretability and scalability?" The main directions will be applying RL for optimizing decision -making for out-of-stock/over-stock, new product recommendations, competitive price setting, al

location with limited supply, and routing. From the question, the following assumptions are mad e by the study: 1) RL-based models will be able to significantly improve principal performance metrics (e.g., cost savings, lead time, and service level) compared to traditional SCM optimizatio n.2) Hybrid RL methods blending deep learning, simulation, and domain expertise can advance s tability and convergence rate and thus make RL tractable for deployment in real SCM applicatio ns.

1.3.Proposed Solution Objectives

This research focuses on applying reinforcement learning (RL) to solve dynamic supply c hain issues in the hope of developing and testing theoretical frameworks rather than creating a sp ecific product. The goal is to explore how RL can improve supply chain decision-making by defi ning it as a Markov Decision Process (MDP) to solve the sequential and stochastic nature of the operations. Key areas of research are the assessment of RL algorithm performance like Q-learnin g, Deep Q-Networks, and Actor-Critic methods in supply chain applications like demand forecas ting, inventory reordering, and dynamic pricing. The study will analyze the trade-off between mo del performance, computational complexity, and applicability in real-world scenarios, enhancing academic literature in terms of the capability of RL for supply chain management.

1.4.Impact of the Proposed Solution

If implemented correctly, the new RL-based approach can alter the functioning of supply cha ins. It could be employed to develop adaptive, data-driven decision-making frameworks that lear n from and respond to changes in the environment, such as changes in the market, delayed suppli es, or disruptions. Such responsiveness not only optimizes for efficiency but can also optimize fo r resilience against global pandemics or geopolitics. Besides, the integration of RL into SCM can support sustainability by maximizing the use of resources, elimination of waste, and minimizatio n of carbon footprints. Ultimately, these intelligent systems could redefine competitive advantag e in worldwide supply chains because they make them autonomous, clear, and open. After the implementation the methodology can be applied to the local warehouse and retailer stores soon for the optimized sales.

2. Review

2.1.Overview of Existing Market

The market itself has developed extremely fast with automation, artificial intelligence, an d analytics and is pushing practices like manufacturing, healthcare, and retail towards intelligent, adaptive supply chain management (SCM) systems (Rolf et al., 2023; Deng et al., 2025). The ma jority of SCM models, however, still operate on the grounds of classical optimization strategies t hat fail under uncertainty or disruption (Kaynov et al., 2023; Oroojlooyjadid et al., 2019). Reinfo rcement Learning (RL) offers an online approach in learning optimal action via experience by act ing and receiving feedback (Giannoccaro & Pontrandolfo, 2002). In spite of computational probl ems (Swazinna et al., 2023), combining model-based and model-free RL, as suggested by Bansal

et al. (2017), would close current gaps and make fully autonomous predictive supply chains a rea lity.

2.2.Gaps in Current Market

Despite advances in analytics and digitalization, RL also encounters significant barriers to applic ation within real-world supply chains. The majority of RL work uses simulated environments suc h as the Beer Game, which causes a "reality gap" where models do not perform under realistic co mplexity (Rolf et al., 2023; Deng et al., 2025). Deep RL's lack of transparency also constrains m anagerial faith, although explainable RL is on the horizon (Chong et al., 2022). Interoperability with existing ERP systems continues to be challenging due to latency and governance issues (Or oojlooyjadid et al., 2019). Moreover, existing RL applications target economic goals, while susta inability and fairness goals remain unexplored in multi-objective supply chain optimization.

3. Proposed Solution

This is a research-oriented project, and it focuses on the application of Reinforcement Learning (RL), specifically Deep Q-Learning, to address dynamic and interdependent problems characteristic of supply chain management. The key issue is to develop an integrated RL platform that can autonomously and effectively make data-driven decisions in many of the key applications with the focus on addressing the out-of-stock/over-stock issue. By representing the supply chain in the form of the Markov Decision Process (MDP), we can encapsulate its stochastic and sequential nature and have the agent learn to generalize optimal policies through interacting with a model of an environment.

3.1 Project Goals

The solution will be evaluated by four interconnected use cases in an integrated decision-making model. The foundation, out-of-stock and over-stock management, dynamically optimizes replenishment timing and quantity to minimize losses and holding costs under demand uncertaint y. Competitive pricing automatically sets prices based on demand, competitor activity, and inven tory, optimizing revenue and margins. A New product recommender system offers personalized promotions to balance exploring new products and exploiting familiar ones. All of these compon ents together constitute an end-to-end RL system that learns, adapts, and optimizes supply chain operations in real time for efficiency, responsiveness, and profitability in dynamic markets.

3.2 System Architecture and Design

The proposed architecture has one Reinforcement Learning (RL) agent at its core, which operates in an end-to-end closed-loop system that learns and evolves continuously with real sale s, inventory, and price information. A simulated environment replicates variation in demand and supplier lead times. Utilizing a Deep Q-Network (DQN), the agent learns to make optimal decisi ons on replenishment, pricing, and supply allocation. The reward function balances service level

s, profitability, and stock control. This integrated framework enables intelligent, data-driven appr oaches to dynamically and effectively optimize multifunctional supply chain operations.

3.3 Technologies and Tools

The system will use PyTorch or TensorFlow to implement the Deep Q-Network (DQN) and its variants, such as Double and Dueling DQN, for improved learning stability. A custom simulation environment will be built using OpenAI Gymnasium, enabling standardized and realistic agent-environment interactions.

3.4 Data Collection and Analysis Methods (if applicable)

Data auditing will evaluate the quality, completeness, and consistency of inputs such as sales records, inventory levels, and purchase orders. Feature engineering will transform raw data into informative state variables, including demand trends, seasonality patterns, and price elasticity indicators. Finally, environment calibration will align the simulation closely with historical data to ensure it realistically mirrors real-world supply chain dynamics before training the RL agent.

3.5 Evaluation Metrics and Success Criteria

The system's performance will be assessed through three metric categories; out-of-stock metrics, business, and algorithmics, with a primary focus on out-of-stock management. Primary metrics will measure how effectively the system balances availability and efficiency. The service level tracks the percentage of demand met, and the rate of stockouts will be minimized to reduce both frequency and duration. Business metrics focus on lowering total supply chain costs and improving profitability through higher Gross Margin Return on Inventory (GMROI). Algorithmic metrics evaluate training stability, ensuring smooth convergence of rewards, and computational efficiency, confirming the model runs within reasonable business timeframes.

4. Expected Outcomes

4.1.Anticipated Deliverables

This research aims to develop an RL-based decision optimization system for supply chain management with a focus on sales and distribution. The system, developed on the basis of a mod ular simulation platform using Markov Decision Processes (MDP), allows continuous interaction between the RL agent and virtual supply chain parameters like demand, transport, and stock. A D eep Q-Network (DQN) model will be used to optimize stock, allocation, and price in the face of uncertainty. There will be an integrated dashboard that compares KPIs such as service levels, sto ckouts, and cost between RL and traditional means. The research concludes with a critical analys is and academic report that is a contribution to research in intelligent supply chain optimization.

4.2.Potential Impact and Applications

The proposed framework is academically, industrially, and societally robust in its contribution. A cademically, it supports RL uptake in supply chain management with a reproducible, theory-to-p ractice framework. Industrially, it enhances autonomous decision-making, with the potential for r eal-time adjustment to demand changes (Khezr & Mishra, 2023). The RL-based framework predi cts more accurately, reduces cost, and optimizes inventory (Oroojlooyjadid et al., 2019). Deep Q-Learning minimizes waste and encourages sustainability (AbdElwahab et al., 2025). It bridges ac ademia and business (Giannoccaro & Pontrandolfo, 2002), facilitating greener and more resource -efficient business, particularly for SMEs. Last but not least, the project views RL as a mechanis m of self-improvement, developing adaptive, intelligent, and sustainable supply chain systems.

5. Ethical Considerations (if applicable)

5.1.Data Privacy and Security

Data protection and privacy are the primary ethical goals for RL-driven supply chain management. The platform relies on private information such as customers' purchase history, suppliers, and prices that, if misused, can take organizations into privacy competitive invasion (Rolf et al., 2023). To ease these threats, the project will hold to data privacy regulations like the General Data Protection Regulation (GDPR) and employ anonymization, encryption, and access control with security (Chong et al., 2022; Kaynov et al., 2023). This will be supported by role-based access documentation and policies. Since RL agents are constantly learning from real-world data, monitoring bias and data drift is essential to offer ethical consistency when it comes to making decisions (Demizu et al., 2023). Constant auditing will ensure that the model is not increasing unnecessary patterns or uncovering sensitive information. This is in accordance with values of responsible AI such as confidentiality, fairness, and stakeholder trust across the data lifecycle (Deng et al., 2025).

5.2.Ethical Implications of the Project

Besides privacy, RL use in supply chain management also has ethical issues with fairness, accountability, and sustainability. RL algorithms maximize objectives like cost or efficiency but potentially at the cost of small suppliers, employee turnover, or locality bias (Oroojlooyjadid et al., 2019). These will be sidestepped by incorporating fairness and sustainability measures into the reward system, balancing profit against environmental and social responsibility (Giannoccaro & Pontrandolfo, 2002). Interpretability will be prioritized to avoid "black-box" behavior and ensure managerial control of decisions (Swazinna et al., 2023). Human monitoring will be at the center of all key processes, with accountability and not automating completely leading to unethical conduct (Gao & Want, 2023). The project also considers environmental and social impacts, such as the consumption of computational power and potential displacement of human activities (Rolf et al., 2023). Overall, by incorporating fairness and transparency into its system, this study encourages responsible innovation such that RL

strengthens and does not degrade equity, trust, and sustainability in international supply chains (Deng et al., 2025).

6. Conclusion

This project proposes a new Reinforcement Learning (RL) method based on Deep Q-Lear ning to transform supply chain management into a strategic adaptive system from a reactive proc ess. It addresses the significant issues of stock imbalances, price pressures, and supply shortage t hrough intelligent decision-making. With real-world customer data and a very scalable simulatio n platform, the RL agent will learn to optimize performance and maximize key business metrics l ike service levels, cost reductions, and top-line growth. Withstanding data complexity and comp utational needs, the project bridges the gap between practice and theory, demonstrating how AI c an make supply chains flexible, efficient, and robust in the modern rapidly changing global mark et.

7. References

[1] B. Rolf, I. Jackson, M. Müller, S. Lang, T. Reggelin, and D. Ivanov, "A review on reinforcement learning algorithms and applications in supply chain management," International Journal of Production Research, vol. 61, no. 20, pp. 7151–7179, 2023, doi: 10.1080/00207543.2022.2140221.

[2] Y. Deng, A. H. F. Chow, Y. Yan, Z. Su, Z. Zhou, and Y.-H. Kuo, "Hierarchical production control and distribution planning under retail uncertainty with reinforcement learning," International Journal of Production Research, vol. 63, no. 12, pp. 4504–4522, 2025, doi: 10.1080/00207543.2025.2452386.

[3] J. W. Chong, W. Kim, and J. S. Hong, "Optimization of apparel supply chain using deep reinforcement learning," IEEE Access, vol. 10, pp. 98743–98759, Sep. 2022, doi: 10.1109/ACCESS.2022.3205720.

[4] I. Kaynov, M. van Knippenberg, V. Menkovski, A. van Breemen, and W. van Jaarsveld, "Deep reinforcement learning for one-warehouse multi-retailer inventory management," International Journal of Production Economics, vol. 267, no. 109088, Nov. 2023, doi: 10.1016/j.ijpe.2023.109088.

[5] A. Oroojlooyjadid, M. R. Nazari, L. V. Snyder, and M. Takáč, "Reinforcement learning algorithm to solve inventory optimization problems," Department of Industrial and Systems Engineering, Lehigh University, Bethlehem, PA, USA, 2019.

[6] I. Giannoccaro and P. Pontrandolfo, “Inventory management in supply chains: a reinforcement learning approach,” International Journal of Production Economics, vol. 78, pp. 153–161, 2002, doi: 10.1016/S0925-5273(01)00159-4.

[7] T. Demizu, Y. Fukazawa, and H. Morita, “Inventory management of new products in retailers using model-based deep reinforcement learning,” Expert Systems with Applications, vol. 229, no. 120256, May 2023, doi: 10.1016/j.eswa.2023.120256.

[8] P. Swazinna, S. Udluft, D. Hein, and T. Runkler, “Comparing model-free and model-based algorithms for offline reinforcement learning,” IFAC-PapersOnLine, vol. 56, no. 2, pp. 3278–3285, 2023, doi: 10.1016/j.ifacol.2023.07.105.

[9] C. Gao and D. Wang, “Comparative study of model-based and model-free reinforcement learning control performance in HVAC systems,” Journal of Building Engineering, vol. 72, no. 108567, May 2023, doi: 10.1016/j.jobe.2023.108567.

[10] S. Bansal, R. Calandra, K. Chua, S. Levine, and C. Tomlin, “MBMF: Model-based priors for model-free reinforcement learning,” arXiv preprint, arXiv:1709.03153, 2017.

[11] M. AbdElwahab, S. ElGazzar, K. Mahhar, and H. Abdel Kader, “Optimizing food SME inventory using a deep reinforcement learning framework,” Journal of Information Systems Engineering and Management, vol. 10, no. 37s, pp. 9–12, 2025, doi: 10.52783/jisem.v10i37s.6374.

[12] L. Khezr and N. Mishra, “A review on reinforcement learning algorithms and applications in supply chain management,” Computers & Industrial Engineering, vol. 185, p. 109564, 2023, doi: 10.1016/j.cie.2023.109564.