

Weronika ŁYSZCZAN, Roger KSIAŻEK, Klaudia KRAWIEC, Piotr ŁEBKOWSKI
AGH Akademia Górniczo-Hutnicza

ZASTOSOWANIE UCZENIA PRZEZ WZMACNIANIE DO WYZNACZANIA WIELKOŚCI ZAMÓWIEŃ W PRZEDSIĘBIORSTWIE

Streszczenie. W przedsiębiorstwie zaobserwowano problem związany z wyznaczaniem wielkości zamówień. Celem badań było opracowanie i ocena metody wyznaczania wielkości zamówień opartej na uczeniu przez wzmocnienie oraz porównanie jej z metodą punktu ponownego zamawiania na podstawie rzeczywistych danych. Uzyskane wyniki wskazują, że zastosowanie uczenia przez wzmocnienie umożliwia poprawę jakości decyzji dotyczących uzupełniania zapasów.

DETERMINATION OF ORDER SIZE IN THE ENTERPRISE USING MACHINE LEARNING

Summary. A problem related to determining order quantities was identified in the company. The aim of the study was to develop and evaluate a reinforcement learning-based method for determining order quantities and to compare it with the re-order point policy using real-world data. The results indicate that reinforcement learning improves the quality of inventory replenishment decisions.

1. Zarządzanie zamówieniami i zapasami

Efektywne zarządzanie zapasami stanowi jeden z kluczowych elementów logistyki przedsiębiorstwa i w istotny sposób wpływa na poziom obsługi klienta, płynność procesów operacyjnych oraz wyniki ekonomiczne organizacji. Podstawowym celem gospodarki magazynowej jest zapewnienie dostępności produktów przy jednoczesnym ograniczaniu kosztów związanych z ich zamawianiem, magazynowaniem oraz utrzymywaniem odpowiedniego poziomu zapasów.

Gospodarka magazynowa obejmuje nie tylko magazyny, infrastrukturę i pracowników, lecz także procesy zachodzące w poszczególnych strefach przedsiębiorstwa [6]. Jej podstawowym procesem jest magazynowanie, rozumiane jako zespół czynności związanych z przyjmowaniem, składowaniem, przechowywaniem, kompletowaniem, przemieszczaniem oraz wydawaniem zapasów [12]. Współczesne magazyny pełnią kluczową rolę w koordynacji podaży i popytu, wspieraniu procesów produkcyjnych oraz działalności marketingowej. Ich zadaniem jest zapewnienie odpowiedniej dostępności produktów przy jednoczesnej minimalizacji kosztów utrzymania zapasów oraz kosztów niedoborów [11].

Utrzymywanie zapasów jest niezbędne ze względu na zmienność popytu, niepewność dostaw, sezonowość sprzedaży oraz konieczność zapewnienia ciągłości procesów produkcyjnych i sprzedażowych. Z drugiej strony nadmierny poziom zapasów

prowadzi do zamrożenia kapitału, wzrostu kosztów magazynowania oraz zwiększenia ryzyka utraty wartości przechowywanych produktów. Zbyt niski poziom zapasów skutkuje natomiast niedoborami, opóźnieniami realizacji zamówień oraz utratą potencjalnej sprzedaży [7].

W literaturze koszty związane z zarządzaniem zapasami najczęściej dzieli się na trzy podstawowe grupy: koszty zamawiania, koszty utrzymywania zapasów oraz koszty niedoboru. Koszty zamawiania obejmują wszystkie działania związane z przygotowaniem i realizacją zamówienia u dostawcy. Koszty utrzymywania zapasów wynikają z konieczności magazynowania produktów, zaangażowania kapitału oraz starzenia się zapasów. Natomiast koszty niedoboru są konsekwencją braku dostępności produktów i obejmują między innymi utracone przychody, przestoje produkcyjne, opóźnienia dostaw oraz pogorszenie poziomu obsługi klienta [13].

Obniżenie całkowitych kosztów gospodarki magazynowej wymaga odpowiedniego doboru strategii uzupełniania zapasów. W praktyce przedsiębiorstw stosowane są zarówno klasyczne modele deterministyczne, jak i coraz częściej metody wykorzystujące analizę danych oraz techniki sztucznej inteligencji wspomagające podejmowanie decyzji.

Sterowanie zapasami polega na podejmowaniu decyzji dotyczących momentu złożenia zamówienia oraz wielkości partii dostawy. Celem tych działań jest zapewnienie wysokiego poziomu dostępności produktów przy możliwie najniższych kosztach funkcjonowania systemu magazynowego.

Jedną z najczęściej stosowanych metod jest metoda punktu ponownego zamawiania (ang. *re-order point policy*, skrót ROP). W podejściu tym zamówienie składane jest w chwili, gdy poziom zapasu osiągnie wcześniej wyznaczoną wartość alarmową. Wielkość zamówienia pozostaje stała, co oznacza, że każdorazowe uzupełnienie zapasu następuje przez dostawę partii o tej samej wielkości [13].

Inną koncepcją w zarządzaniu zapasami jest ekonomiczna wielkość zamówienia (ang. *Economic Order Quantity*, skrót EOQ), zaproponowana przez Harrisa już w 1913 roku, umożliwia wyznaczenie takiej wielkości partii dostawy, która minimalizuje sumę kosztów zamawiania i utrzymywania zapasów [5]. Model zakłada między innymi stały popyt, niezmienny czas dostawy, brak rabatów ilościowych oraz możliwość realizacji zamówienia w dowolnym momencie [4].

Alternatywnym rozwiązaniem jest model stałego cyklu zamawiania (ang. *re-order cycle point*, skrót ROC), w którym kontrola zapasu odbywa się w ustalonych odstępach czasu, natomiast wielkość zamówienia zależy od aktualnego poziomu zapasu oraz ilości towaru znajdującego się w drodze [13]. W praktyce przemysłowej spotyka się również rozwiązania wykorzystujące poziomy minimalne i maksymalne wyznaczone na podstawie prognoz popytu. Po przekroczeniu poziomu MIN danego produktu, zamawiana jest wielkość, którą stanowi różnicę między stanem MAX, a aktualnym stanem magazynu [3].

Pomimo swojej prostoty klasyczne modele zarządzania zapasami opierają się na szeregu założeń upraszczających rzeczywistość. Współczesne przedsiębiorstwa funkcjonują jednak w warunkach zmiennego popytu, niepewności dostaw oraz dużej liczby indeksów magazynowych, przez co stosowanie wyłącznie modeli deterministycznych nie zawsze prowadzi do uzyskania bardzo dobrych decyzji.

2. Opis analizowanego przedsiębiorstwa i problemu decyzyjnego

Badania przeprowadzono w przedsiębiorstwie zajmującym się dystrybucją rozwiązań z zakresu automatyki przemysłowej. Firma współpracuje z wieloma zagranicznymi dostawcami, a część oferowanych produktów charakteryzuje się długim czasem realizacji zamówień, który w niektórych przypadkach przekracza trzy miesiące. Jednocześnie przedsiębiorstwo utrzymuje szeroki asortyment produktów magazynowych, co znacząco utrudnia podejmowanie decyzji dotyczących uzupełniania zapasów.

Obecnie decyzje o wielkości zamówień podejmowane są przez kierownika importu na podstawie analizy historycznej sprzedaży, zdefiniowanych poziomów minimalnych i maksymalnych oraz własnego doświadczenia. Takie podejście, mimo wieloletniej praktyki osoby odpowiedzialnej za zamówienia, jest w znacznym stopniu uzależnione od subiektywnej oceny i nie uwzględnia wszystkich czynników wpływających na przyszłe zapotrzebowanie.

Konsekwencją obecnego sposobu podejmowania decyzji jest okresowe występowanie zarówno nadmiernych zapasów, powodujących wzrost kosztów magazynowania i zamrożenie kapitału, jak również niedoborów prowadzących do utraty sprzedaży oraz pogorszenia poziomu obsługi klientów. Problem ten staje się szczególnie istotny w przypadku produktów importowanych, dla których długi czas realizacji zamówienia znacząco ogranicza możliwość szybkiego reagowania na zmieniający się popyt.

Z tego względu zasadne wydaje się zastosowanie metod wspomagania decyzji, które potrafią uwzględniać zmienność środowiska oraz uczyć się optymalnych strategii zamawiania na podstawie danych historycznych. W niniejszej pracy zaproponowano wykorzystanie algorytmu uczenia przez wzmacnianie (ang. *Reinforcement Learning*, skrót RL), którego skuteczność została porównana z klasyczną metodą punktu ponownego zamawiania. Celem przeprowadzonych badań było określenie, czy zastosowanie uczenia ze wzmacnianiem pozwala ograniczyć całkowite koszty gospodarki magazynowej oraz zmniejszyć poziom niezaspokojonego popytu w porównaniu z klasycznym podejściem deterministycznym.

3. Uczenie przez wzmacnianie w zarządzaniu zapasami

Uczenie maszynowe (ang. *Machine Learning*, skrót ML) jest jedną z najdynamiczniej rozwijających się dziedzin sztucznej inteligencji, której celem jest tworzenie modeli zdolnych do automatycznego pozyskiwania wiedzy z danych oraz wykorzystywania jej do podejmowania decyzji lub prognozowania przyszłych zdarzeń. Klasyczna definicja zaproponowana przez Toma Mitchella określa uczenie maszynowe jako proces, w którym program komputerowy poprawia swoje działanie w realizacji zadania T na podstawie doświadczenia E , mierzonego określoną miarą jakości P [8]. Oznacza to, że zamiast implementowania sztywnych reguł decyzyjnych, model samodzielnie identyfikuje zależności występujące w danych uczących i wykorzystuje je podczas analizy nowych obserwacji.

Współczesne systemy uczenia maszynowego znajdują zastosowanie wszędzie tam, gdzie budowa klasycznych algorytmów opartych na regułach okazuje się niewystarczająca lub zbyt kosztowna. Dotyczy to przede wszystkim problemów charakteryzujących się dużą zmiennością środowiska, znaczną liczbą zmiennych opisujących proces

oraz koniecznością ciągłego dostosowywania podejmowanych decyzji do nowych danych [9]. Dzięki temu uczenie maszynowe jest obecnie szeroko wykorzystywane między innymi w analizie danych, prognozowaniu popytu, systemach rekomendacyjnych, diagnostyce medycznej, wykrywaniu nadużyć finansowych oraz optymalizacji procesów logistycznych i produkcyjnych [10].

Najczęściej wyróżnia się cztery podstawowe paradygmaty uczenia maszynowego: uczenie nadzorowane, nienadzorowane, półnadzorowane oraz uczenie przez wzmacnianie. W uczeniu nadzorowanym model budowany jest na podstawie oznaczonych przykładów uczących, dzięki czemu możliwe jest rozwiązywanie problemów klasyfikacji i regresji. Uczenie nienadzorowane wykorzystuje wyłącznie dane wejściowe i służy przede wszystkim do odkrywania struktur, grupowania obserwacji oraz wykrywania anomalii. Z kolei uczenie półnadzorowane stanowi połączenie obu podejść i znajduje zastosowanie w sytuacjach, gdy jedynie niewielka część danych posiada etykiety.

Niezależnie od zastosowanego paradygmatu skuteczność modeli uczenia maszynowego jest w dużej mierze uzależniona od jakości danych. Dane uczące powinny być reprezentatywne dla rzeczywistych warunków pracy modelu, odpowiednio oczyszczone oraz zawierać cechy opisujące analizowane zjawisko. Do najczęściej spotykanych problemów należą niedobór danych, ich niska jakość, obecność obserwacji odstających oraz zjawiska przetrenowania i niedotrenowania modeli, które prowadzą odpowiednio do utraty zdolności uogólniania lub niewystarczającego odwzorowania zależności występujących w danych [9].

Projektowanie systemu uczenia maszynowego ma charakter iteracyjny i obejmuje przygotowanie danych, wybór reprezentacji cech, budowę modelu, jego walidację, wdrożenie oraz ciągle monitorowanie jakości działania w środowisku produkcyjnym [10]. W praktyce przemysłowej równie istotna jak sama konstrukcja modelu jest jego zdolność do adaptacji do zmieniających się warunków funkcjonowania przedsiębiorstwa. W wielu zastosowaniach, zwłaszcza związanych z logistyką i zarządzaniem zapasami, dane napływają w sposób ciągły, a charakter popytu zmienia się w czasie, dlatego modele powinny umożliwiać aktualizację wiedzy wraz z pojawianiem się nowych obserwacji.

Spośród wszystkich paradygmatów uczenia maszynowego szczególną uwagę zwraca uczenie przez wzmacnianie, którego celem nie jest jedynie przewidywanie wartości lub klasyfikacja danych, lecz bezpośrednie wyznaczanie strategii podejmowania decyzji maksymalizujących długoterminową wartość określonej funkcji celu. Z tego względu metoda ta znajduje coraz szersze zastosowanie w problemach optymalizacyjnych i systemach wspomagania decyzji, w których kolejne decyzje wpływają na przyszłe stany analizowanego procesu [9].

Uczenie przez wzmacnianie (ang. *Reinforcement Learning*, skrót RL) opiera się na interakcji agenta ze środowiskiem. W odróżnieniu od uczenia nadzorowanego agent nie otrzymuje poprawnych odpowiedzi dla poszczególnych przykładów uczących. Zamiast tego uczy się na podstawie konsekwencji własnych działań, otrzymując nagrody lub kary. Proces ten prowadzi do stopniowego doskonalenia strategii działania (ang. policy) umożliwiającej osiągnięcie coraz lepszych wyników.

Jednym z algorytmów uczenia przez wzmacnianie jest proces decyzyjny Markowa. Model ten definiuje zbiór stanów środowiska, możliwe działania agenta, funkcję nagrody oraz prawdopodobieństwa przejść pomiędzy kolejnymi stanami. Zadaniem

agenta jest znalezienie takiej polityki działania, która maksymalizuje zdyskontowaną sumę przyszłych nagród. Bellman odkrył sposób jak oszacować optymalną wartość stanu s , która jest określana wyrażeniem:

$$V^*(s) = \max_a \sum_{s'} T(s, a, s') [R(s, a, s') + \gamma V^*(s')] \text{ dla wszystkich } s, \quad (1)$$

gdzie $T(s, a, s')$ oznacza prawdopodobieństwo przejścia pomiędzy stanami, $R(s, a, s')$ określa otrzymywaną nagrodę, natomiast γ jest współczynnikiem dyskontowym określającym znaczenie przyszłych nagród.

W praktycznych zastosowaniach znacznie częściej wykorzystywana jest funkcja wartości stanu - czynności, nazywana Q-wartością, opisująca oczekiwaną wartość wykonania określonej akcji w danym stanie. Pozwala ona agentowi bezpośrednio wybierać działanie prowadzące do maksymalizacji długoterminowej nagrody zgodnie z zależnością:

$$\pi^*(s) = \arg_a \max Q^*(s, a), \quad (2)$$

gdzie π^* oznacza politykę optymalną [9]. Współczesne algorytmy uczenia przez wzmacnianie wykorzystują aproksymację funkcji wartości za pomocą sieci neuronowych, dzięki czemu możliwe jest rozwiązywanie problemów charakteryzujących się bardzo dużą liczbą stanów i decyzji.

W ostatnich latach uczenie przez wzmacnianie znalazło zastosowanie w wielu obszarach wymagających sekwencyjnego podejmowania decyzji, takich jak sterowanie robotami, systemy autonomiczne, zarządzanie energią, planowanie produkcji, optymalizacja procesów logistycznych oraz gospodarka magazynowa [10]. W przeciwieństwie do klasycznych metod sterowania zasobami agent RL nie wykorzystuje wcześniej zdefiniowanych reguł decyzyjnych. Samodzielnie uczy się polityki zamawiania na podstawie obserwacji środowiska, uwzględniając jednocześnie wzajemne zależności pomiędzy poziomem zapasów, kosztami magazynowania, kosztami zamówień oraz ryzykiem wystąpienia niedoborów.

Implementację oraz eksperymenty obliczeniowe przeprowadzono w języku Python, który obecnie stanowi jeden z najczęściej wykorzystywanych języków programowania w projektach związanych z analizą danych i sztuczną inteligencją [2]. Popularność tego środowiska wynika przede wszystkim z bogatego ekosystemu bibliotek wspomagających przetwarzanie danych, budowę modeli uczenia maszynowego oraz realizację eksperymentów obliczeniowych. Wśród najczęściej wykorzystywanych narzędzi znajdują się między innymi Scikit-learn, TensorFlow oraz PyTorch, umożliwiające implementację zarówno klasycznych algorytmów uczenia maszynowego, jak i zaawansowanych modeli głębokiego uczenia [1].

4. Wyniki eksperymentów obliczeniowych

Ocena zaproponowanych metod została przeprowadzona z wykorzystaniem rzeczywistych danych przedsiębiorstwa zajmującego się dystrybucją elementów automatyki przemysłowej. Dane obejmowały okres od 1 stycznia 2022 r. do 30 września 2023 r., co odpowiada 638 dniom obserwacji. Udostępniony zbiór zawierał informacje dotyczą-

ce sprzedaży, zamówień klientów, zamówień składanych u dostawców oraz terminów realizacji dostaw.

Do badań wybrano pięć produktów spośród około 5000 indeksów magazynowych dostępnych w ofercie przedsiębiorstwie. Wybrane produkty charakteryzowały się wysoką rotacją magazynową oraz dużym znaczeniem sprzedażowym, dzięki czemu stanowiły reprezentatywną grupę do oceny metod zarządzania zapasami. Wszystkie analizowane produkty pochodziły od tego samego dostawcy, dla którego przyjęto średni czas realizacji zamówienia równy 12 dni.

Przed rozpoczęciem eksperymentów dane zostały poddane procesowi przygotowania. Usunięto pojedyncze, nietypowe zamówienia o bardzo dużej liczbie, które mogły zaburzać estymację rozkładów popytu. Na podstawie oczyszczonych danych wyznaczono dyskretne empiryczne rozkłady prawdopodobieństwa dziennego popytu dla każdego produktu. W większości dni nie występowała sprzedaż, natomiast sporadycznie pojawiały się pojedyncze zamówienia o znacznie większej liczbie. Taka charakterystyka popytu jest typowa dla przedsiębiorstw dystrybucyjnych oferujących szeroki asortyment produktów specjalistycznych.

Na potrzeby eksperymentów przyjęto jednakowy koszt złożenia zamówienia wynoszący 300 zł dla wszystkich produktów oraz koszt utrzymywania zapasu odpowiadający 15% wartości produktu w skali roku. Parametry te zostały oszacowane na podstawie danych przedsiębiorstwa obejmujących koszty pracy, obsługi administracyjnej oraz magazynowania.

4.1. Rozwiązania rzeczywiste jako punkt odniesienia

Pierwszym etapem badań była analiza rzeczywistego sposobu zarządzania zapasami stosowanego w przedsiębiorstwie. Dla każdego produktu wyznaczono całkowite koszty obejmujące koszty zamawiania, koszty utrzymywania zapasu oraz koszty wynikające z niezaspokojonego popytu. Dodatkowo obliczono średni poziom straconego popytu stanowiący drugie kryterium oceny jakości rozwiązania.

Uzyskane wyniki wykazały znaczące różnice pomiędzy analizowanymi produktami. Największe całkowite koszty odnotowano dla produktu nr 1, co wynikało przede wszystkim z wysokiej wartości jednostkowej oraz wysokich kosztów utrzymywania zapasu. Z kolei największy udział niezaspokojonego popytu wystąpił dla produktu nr 5, co wskazuje na niedostateczne zabezpieczenie zapasów dla tego asortymentu. Otrzymane wyniki potwierdzają, że obecnie stosowana polityka zamawiania nie zapewnia jednoczesnej minimalizacji kosztów oraz utrzymania wysokiego poziomu obsługi klienta.

Tabela 1 przedstawia zestawienie wyników uzyskanych dla rzeczywistej polityki zarządzania zapasami.

4.2. Zastosowanie polityki ROP

Jako punkt odniesienia dla zaproponowanego rozwiązania wykorzystano klasyczną metodę punktu ponownego zamawiania (ROP). Dla każdego produktu przeprowadzono symulację Monte Carlo (skrót MC), w której parametry polityki (Q , s) wyznaczano metodą przeszukiwania siatki. Dla każdej analizowanej kombinacji parametrów wykonano 10 000 niezależnych przebiegów symulacyjnych, co pozwoliło ograniczyć wpływ losowości generowanego popytu na otrzymywane wyniki.

Kryterium wyboru rozwiązania stanowiła minimalizacja całkowitych kosztów przy jednoczesnym zachowaniu średniego poziomu straconego popytu poniżej 5%. Oce-

Tabela 1

Suma całkowitych kosztów poniesionych przez przedsiębiorstwo w zależności od danego produktu oraz średni stracony popyt dla okresu 638 dni.

Nr produktu	Suma całkowitych kosztów [zł]	Średni stracony popyt [%]
Produkt nr 1	175 539,46 zł	10,58%
Produkt nr 2	64 851,41 zł	13,40%
Produkt nr 3	56 066,45 zł	16,78%
Produkt nr 4	29 795,58 zł	0,00%
Produkt nr 5	29 965,72 zł	17,12%

nie podlegały wyłącznie rozwiązaniu spełniające to ograniczenie.

Przeprowadzone eksperymenty wykazały, że optymalne wartości parametrów polityki ROP różniły się pomiędzy analizowanymi produktami. Jednocześnie dla wszystkich przypadków możliwe było uzyskanie znacznej redukcji całkowitych kosztów w porównaniu z rozwiązaniem rzeczywistym przedsiębiorstwa. Największe korzyści uzyskano dla produktów charakteryzujących się wysoką wartością jednostkową oraz dużą zmiennością popytu.

Na rysunku 1 przedstawiono wyniki przeprowadzonej symulacji Monte Carlo dla produktu nr 1. Wyznaczone całkowite koszty nie zmieniają się liniowo, ale wykazują wyraźne cykle. Gdy wielkość zamówienia jest na poziomie 10 lub 20 sztuk widać, że wyznaczone koszty są bardzo wysokie w porównaniu do innych zastosowanych polityk, wyjątkiem jest rozwiązanie $Q = 20$, $s = 10$, gdzie występuje znaczący spadek kosztów. Następnie można zauważyć pewną zależność dla pozostałych wartości Q . Gdy punkt ponownego zamawiania jest na poziomie od 10 do 30 sztuk koszty maleją, natomiast później zaczynają rosnąć. Ponadto wraz ze wzrostem wartości Q , różnice wyznaczonych całkowitych kosztów pomiędzy zastosowanymi politykami maleją. Najniższe całkowite koszty o wartości 74 774,10 zł otrzymano dla rozwiązania $Q = 40$, $s = 30$.

Tabela 2 zestawiono wyniki, czyli najniższe całkowite koszty oraz średni stracony popyt dla każdego produktu, które uzyskano poprzez wykonanie symulacji Monte Carlo dla rozwiązań wyznaczonych przy zastosowaniu metody ROP.

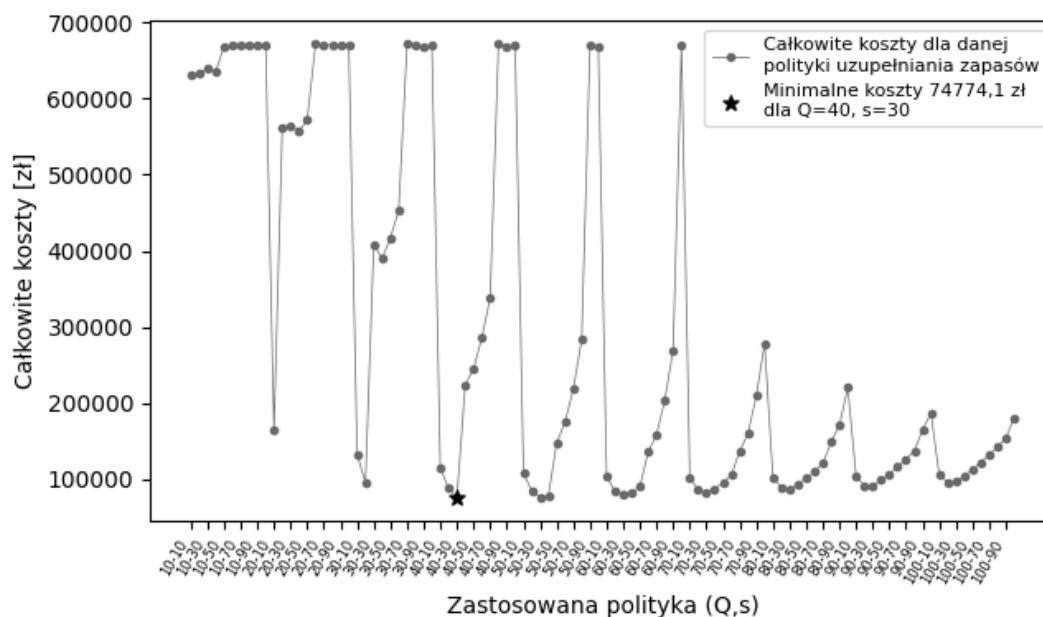
Tabela 2

Najniższe całkowite koszty oraz średni stracony popyt otrzymany po przeprowadzeniu symulacji MC dla każdego produktu.

Nr produktu	Całkowite koszty [zł]	Średni stracony popyt [%]
Produkt nr 1	74 774,10 zł	3,84%
Produkt nr 2	25 188,60 zł	2,22%
Produkt nr 3	11 402,50 zł	2,09%
Produkt nr 4	11 350,20 zł	3,95%
Produkt nr 5	5 741,60 zł	4,15%

4.3. Zastosowanie uczenia przez wzmacnianie

Ostatni etap badań obejmował ocenę zaproponowanej metody wykorzystującej uczenie przez wzmacnianie. Agent RL uczył się samodzielnie podejmowania decyzji



Rys. 1. Wyniki przeprowadzonej symulacji MC dla produktu nr 1.

dotyczących wielkości zamówień poprzez interakcję ze środowiskiem symulacyjnym odwzorowującym proces funkcjonowania magazynu. Funkcja nagrody została zdefiniowana w taki sposób, aby minimalizować całkowite koszty przy jednoczesnym ograniczeniu średniego straconego popytu.

W przeprowadzonych badaniach zastosowano klasyczny algorytm Q-learning. Agent w kolejnych dniach obserwował stan środowiska opisujący sytuację magazynową (m.in. poziom zapasu oraz zamówienia w realizacji), a następnie podejmował decyzję dotyczącą wielkości zamówienia. Po wykonaniu akcji środowisko było aktualizowane poprzez uwzględnienie losowo generowanego popytu oraz dostaw, a agent otrzymywał nagrodę zależną od całkowitych kosztów obejmujących koszty zamawiania, utrzymywania zapasów i straconego popytu. Na podstawie otrzymanej nagrody aktualizowana była tablica wartości $Q(s, a)$, dzięki czemu agent stopniowo uczył się polityki zamawiania minimalizującej długookresowe koszty.

Przestrzeń akcji stanowiły dopuszczalne wielkości zamówień, natomiast proces uczenia prowadzono w środowisku symulacyjnym odwzorowującym rzeczywisty proces funkcjonowania magazynu.

Uzyskane wyniki porównano zarówno z rzeczywistą polityką stosowaną przez przedsiębiorstwo, jak również z rozwiązaniami wyznaczonymi metodą ROP. Dla wszystkich analizowanych produktów metoda wykorzystująca uczenie przez wzmacnianie umożliwiła uzyskanie stabilnych polityk zamawiania zapewniających niski poziom kosztów oraz ograniczenie średniego straconego popytu.

Największą zaletą proponowanego podejścia okazała się możliwość automatycznego dostosowywania decyzji do zmieniających się warunków popytu bez konieczności ręcznego doboru parametrów polityki zamawiania. W przeciwieństwie do klasycznej metody ROP, agent nie wymaga wcześniejszego określenia punktu ponownego zamawiania ani wielkości partii dostawy, lecz wyznacza strategię działania bezpośrednio na podstawie doświadczeń zdobytych podczas procesu uczenia.

W tabeli 3 zestawiono całkowite koszty uzyskane metodą ROP oraz metodą uczenia przez wzmacnianie ze znanymi kosztami rzeczywistymi. Analizując otrzymane wyniki, można zauważyć, że metoda ROP znacząco obniża koszty w porównaniu z rzeczywistymi kosztami dla wszystkich produktów, osiągając największą redukcję dla produktu nr 1, czyli zmniejszając koszty aż o 43%. Metoda uczenia ze wzmacnianiem również obniża całkowite koszty w porównaniu do kosztów rzeczywistych, choć nie tak bardzo jak metoda ROP. Jedynie dla produktu nr 2, metoda ta wygenerowała niższe całkowite koszty niż metoda ROP.

Zestawienie średniego straconego popytu uzyskanego metodą ROP oraz metodą uczenia ze wzmacnianiem z wartościami rzeczywistymi przedstawiono w tabeli nr 4. Zarówno metoda ROP, jak i metoda uczenia ze wzmacnianiem znacznie obniżają średni stracony popyt w porównaniu z rzeczywistymi wartościami. W tym przypadku metoda uczenia ze wzmacnianiem jest znacznie bardziej efektywna w redukcji straconego popytu, osiągając niemal zawsze zerowe wartości dla wszystkich produktów. Jest to szczególnie widoczne w przypadku produktu nr 3, gdzie stracony popyt zmniejsza się z 16,78% do zaledwie 0,14%.

Tabela 3

Zestawienie całkowitych kosztów otrzymanych metodą ROP oraz metodą uczenia przez wzmacnianie z kosztami rzeczywistymi [zł].

Nr produktu	Koszty rzeczywiste	Metoda ROP	Metoda uczenia ze wzmacnianiem
Produkt nr 1	175 539,46 zł	74 774,10 zł	97 650,58 zł
Produkt nr 2	64 851,41 zł	25 188,60 zł	22 912,44 zł
Produkt nr 3	56 066,45 zł	11 402,50 zł	29 762,80 zł
Produkt nr 4	29 795,58 zł	11 350,20 zł	21 992,36 zł
Produkt nr 5	29 965,72 zł	5 741,60 zł	11 169,80 zł

Tabela 4

Zestawienie średniego straconego popytu otrzymanego metodą ROP oraz metodą uczenia ze wzmacnianiem z wartościami rzeczywistymi [%].

Nr produktu	Rzeczywisty średni stracony popyt	Metoda ROP	Metoda uczenia ze wzmacnianiem
Produkt nr 1	10,58%	3,84%	0,19%
Produkt nr 2	13,40%	2,22%	0,14%
Produkt nr 3	16,78%	2,09%	0,14%
Produkt nr 4	0%	3,95%	0,09%
Produkt nr 5	17,12%	4,15%	0,12%

Podsumowanie i wnioski

W artykule przedstawiono metodę wspomaganie decyzji dotyczących uzupełniania zapasów z wykorzystaniem uczenia przez wzmacnianie. Badania przeprowadzono na rzeczywistych danych przedsiębiorstwa zajmującego się dystrybucją elementów automatyki przemysłowej, w którym zaobserwowano trudności związane z określaniem wielkości zamówień u dostawców. Konsekwencją stosowanej polityki były okresowo

występujące nadmierne zapasy oraz niedobory produktów, prowadzące odpowiednio do wzrostu kosztów magazynowania oraz utraty części sprzedaży.

Skuteczność zaproponowanej metody oceniono poprzez porównanie z rzeczywistą polityką zamawiania przedsiębiorstwa oraz klasyczną metodą punktu ponownego zamawiania (ROP). Jako kryteria oceny przyjęto całkowite koszty gospodarki zapasami oraz średni poziom niezaspokojonego popytu.

Zaproponowana metoda wykorzystująca uczenie przez wzmacnianie pozwoliła znacząco ograniczyć poziom niezaspokojonego popytu. Klasyczna metoda ROP uzyskiwała natomiast niższe koszty całkowite dla większości analizowanych produktów. Oznacza to, że wybór metody zależy od przyjętego kryterium optymalizacji.

Uyskane rezultaty wskazują, że uczenie przez wzmacnianie może stanowić skuteczne narzędzie wspomagające decyzje dotyczące zarządzania zapasami, szczególnie w środowiskach charakteryzujących się zmiennym i trudnym do przewidzenia popytem. W przeciwieństwie do klasycznych metod deterministycznych, agent RL uczy się strategii zamawiania bez konieczności ręcznego doboru parametrów polityki, dzięki czemu może lepiej adaptować się do zmieniających się warunków funkcjonowania przedsiębiorstwa.

Przedstawione badania stanowią punkt wyjścia do dalszych prac nad zastosowaniem metod uczenia maszynowego w zarządzaniu zapasami. W kolejnych etapach planowane jest rozszerzenie modelu o większą liczbę produktów, uwzględnienie wielu dostawców oraz zmiennych czasów realizacji zamówień.

Badania zostały sfinansowane z subwencji oraz ze środków programu „Inicjatywa Doskonałości – Uczelnia Badawcza” dla AGH.

LITERATURA

1. 6 narzędzi do Machine Learning, <https://bulldogjob.pl/readme/6-narzedzi-do-machine-learning-w-pythonie>. Dostęp: 2026-06-26.
2. Co to jest język Python? Do czego służy i jak zacząć w nim programować?, <https://expose.pl/co-to-jest-jezyk-python-do-czego-sluzy-i-jak-zaczac-w-nim-programowac/>. Dostęp: 2026-06-26.
3. Metody zarządzania zaopatrzeniem, <https://prologistica.pl/bazawiedzy/optimalizacja-zapasow-magazynowych/metody-zarzadzania-zaopatrzeniem.html>. Dostęp: 2026-06-26.
4. System EOQ lub model magazynowy Wilsona: Zarządzanie zapasami, <https://www.ar-racking.com/pl/blog/system-eoq-lub-model-magazynowy-wilsona-zarzadzanie-zapasami/>. Dostęp: 2026-06-26.
5. Erlenkotter D.: Ford Whitman Harris and the Economic Order Quantity Model, *Operations Research*, 1990, vol. 38, p. 937–946.
6. Galińska B.: *Gospodarka magazynowa*. Difin, Warszawa 2016.
7. Gwynne R.: *Zarządzanie logistyką magazynową*. Wydawnictwo Naukowe PWN, Warszawa 2016.

8. Géron A.: Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow 3e: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media, 2017.
9. Géron A.: Uczenie maszynowe z użyciem Scikit-Learn i TensorFlow. Helion SA, 2020.
10. Hueyn C.: Jak projektować systemy uczenia maszynowego. Iteracyjne tworzenie aplikacji gotowych do pracy. Helion SA, 2023.
11. Korzeniowski A., Weselik A., Skowroński Z., Kaczmarek M.: Zarządzanie logistyką magazynową. Polskie Wydawnictwo Ekonomiczne, Warszawa 1997.
12. Niemczyk A.: Zapasy i magazynowanie tom II. Magazynowanie. Instytut Logistyki i Magazynowania, Poznań 2008.
13. Pisz I., Sęk T., Zielecki W.: Logistyka w przedsiębiorstwie. Polskie Wydawnictwo Ekonomiczne, Warszawa 2013.